AD-A146 843

DTIC FILE COPY

Research Note 84-127

COGNITIVE PREDICTORS OF TANK COMMANDER PERFORMANCE

Jimmie H. Johnson
Essex Corporation

Marshall B. Jones
Hershey Medical Center

Robert S. Kennedy
Essex Corporation

Donald F. Haggard, Contracting Officer's Representative

Submitted by

Donald F. Haggard, Chief
ARI FIELD UNIT AT FORT KNOX, KENTUCKY

and

Harold F. O'Neil, Jr., Director
TRAINING RESEARCH LABORATORY

U. S. Army

Research Institute for the Behavioral and Social Sciences

October 1984

Approved for public release; distribution unlimited.

84 10 22 019

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER Research Note 84-127	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Cognitive Predictors of Tank Commander Performance		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Jimmie H. Johnson, Essex Corporation; Marshall B. Jones, Hershey Medical Center; Robert S. Kennedy, Essex Corporation		8. CONTRACT OR GRANT NUMBER(s) MDA903-82-C-0176
9. PERFORMING ORGANIZATION NAME AND ADDRESS Essex Corporation 333 North Fairfax Street Alexandria, Virginia 22314		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q263743A794
11. CONTROLLING OFFICE NAME AND ADDRESS US Army Research Institute 5001 Eisenhower Ave. Alexandria, VA 22333		12. REPORT DATE October 1984
		13. NUMBER OF PAGES 106
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Job Sample
Selection
Task Analysis
Video Tests

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of the investigation was to determine which students were most likely to be successful tank commanders (TCs). The following tests were administered: five paper-and-pencil tests; two commercial video games and; a microcomputer based video task called Tank Commander, developed to assess performance on job samples identified as being frequently failed by TCs in field exercises. Two groups were tested.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Forty young officer students were administered the Tank Commander test, only. Significant canonical correlations were found between class standing and performance on the Tank Commander test, however, the results were in the unexpected direction, in that students who did well on the Tank Commander test were ranked low in the class.

Thirty enlisted TC students (three classes) were administered all eight of the tests. Criterion measures were supervisory and peer ratings. Significant correlations were obtained between performance on Tank Commander and the criteria, however, the results were equivocal: two of the classes performed in the expected direction (good scorers on the test received high criterion ratings) while the other class performed in the opposite direction. Two of the paper-and-pencil tests combined (Paper Folding and Map Planning) yielded R=.6 with the criterion.

COGNITIVE PREDICTORS OF TANK COMMANDER PERFORMANCE

Final Report

Jimmie H. Johnson

Essex Corporation

Marshall B. Jones, Ph.D.

Hershey Medical Center

Robert S. Kennedy, Ph.D.

Essex Corporation

Prepared For:

U.S. Army Research Institute for the Behavioral

and Social Sciences

Fort Knox Field Unit

Fort Knox, Kentucky 40121

Table of Contents

Table of Contents (Continued)

List of Tables

List of Tables (Continued)

COGNITIVE PREDICTORS OF TANK COMMANDER PERFORMANCE

Purpose of the Research

The tank commander (TC) is responsible for an expensive, complex, lethal weapon system. Effective TC performance is predicated upon adequate skills, knowledge, motivation and extensive training. Selection of appropriate personnel for TC duties is essential if the full potential of armored weapon systems is to be realized in a cost-effective manner. This must be accomplished within the military framework which normally gives most responsibility to those with most rank. The problem then is to know, from all of the pool of applicants who presumably have at least minimum requisite skills, which will make the best TCs. The purpose of the present research was to identify and develop predictors of TC performance in the new armor systems.

The Prediction Problem

Crew performance is a vital ingredient in realizing the full potential of the hardware and software capabilities of sophisticated, technologically advanced, armor weapon systems. On the battlefield of the future, the role of the TC can best be thought of as a system or subsystem manager. The well-rounded TC will be responsible for the operation, maintenance, and crew training of the system. In his description of the Basic Noncommissioned Officer Course (BNCOC), General Wagner (1981) points out that future TCs will have to be more than just weapons and equipment experts, they will have to be highly trained vehicle commanders and trainers. The flexibility afforded by the sophisticated weaponry and maneuverability of advanced armor vehicles, particularly the M1, means that the TC will be faced with decision-making and choice behavior that did not exist in older, slower systems that could operate effectively in a lock-step, do-as-you-are-directed manner.

Goldin and Thorndyke (1981b) stated that military operations involving learning an unfamiliar region through navigation or from a map require skilled performance of spatial tasks. Goldin and Thorndyke (1981a) compared good and poor cognitive mappers on a variety of spatial knowledge acquisition and judgment tasks and reported that good mappers excel in learning a novel environment, learning maps, and making spatial judgments based on memorized maps, but found no differences between good and poor mappers on using maps, map interpretation, and navigation tasks. Thorndyke and Goldin (1981) assessed verbal comprehension, visualization ability, spatial orientation, visual memory and field dependence, with paper-and-pencil aptitude tests, as well as other techniques to assess visual/verbal processing style, exploratory motivation, motivation to consciously control learning, disorientation anxiety and map-using motivation. The results indicate that only spatial abilities distinguish good from poor mappers. The extent to which spatial abilities are required to be a good TC has not been determined, but it is one of the variables that should be considered in the search for predictors of TC performance.

Paper-and-Pencil Tests

Eaton, Bessemer, and Kristiansen (1979) found little evidence to support the use of paper-and-pencil tests as predictors of performance for tank gunners or drivers. Although the results of Phase I of their study were encouraging, the results were not confirmed in Phases II and III. Phase III included TCs as subjects of the study, with gunnery performance as the criterion. Criteria for the three phases changed, in varying degrees, for each phase, therefore, it is difficult to ascertain whether or not criterion unreliability was, or was not, a contributor to the lack of effectiveness of the predictors. It was reported that the correlation between night and day gunnery scores was weak, or nonexistent. One would think that good gunners or TCs would do well both day and night, while poor gunners would not do well in either case. The lack of a correlation between day and night performance may have been due to unreliable criteria. Although reliability coefficients were not reported for the predictors, a test-retest of the ten predictor variables for TC and gunner, administered two months apart to determine if environmental or motivational factors caused differences, resulted in no significant differences. Although some significant correlations were reported, the size of the coefficients were so small (mostly below .30 for zero order correlations and rarely above .50 for multiple regressions) that not enough variance was accounted for to make the tests of practical importance. The authors concluded that a job or task sample approach might be more effective in predicting performance. The failure of the Armed Services Vocational Aptitude Battery (ASVAB) paper-and-pencil tests to predict gunner performance was confirmed by Eaton, Johnson, and Black (1980).

Paper-and-pencil tests as selection devices have exhibited much success in the several occupations for which they have been employed in military jobs (Imhoff & Levine, 1981), but they have not been used much in armor (Eaton, 1978). Some findings have been reported (Greenstein & Hughes, 1977; Thomas & Sternberg, 1964), but those outcomes once empirically validated, have not fared well on replication (Eaton, 1978; Eaton, Bessemer, & Kristiansen, 1979). The obvious interpretation of these outcomes is that paper-and-pencil tests do not tap factors necessary for tank crew success. However, alternative explanations include the possibility that either the potential predictors or the criteria were unstable or unreliable. Any of these possibilities could so reduce power that sample sizes which are commonly used ($N = 50$) would require correlations between the predictor and the criterion to be greater than $r = .40$ to be significant at the $P = .05$ level of significance. The lack of cross validated predictive validity of the paper-and-pencil tests for armor, plus the lack of favor in which paper-and-pencil tests have come to be held in recent years (Glaser & Bond, 1981) have provided the impetus to rediscover apparatus tests.

Microprocessor-Based Tests

Video games have potential as predictors because they have been shown to have good metric properties and because microprocessors permit avoidance of many of the difficulties experienced by the apparatus experimenters; specifically, break down, malfunction, alignment, and calibration problems.

During the Second World War apparatus, primarily perceptual and psychomotor tasks were shown to have excellent predictive validity for the pilot training criterion and other military criteria (Melton, 1947). For practical reasons these tests are no longer used by any of the military services for selection or classification purposes. Video games, the hope is, may have similar predictive properties to the old apparatus tests. If so, that fact would make possible much better prediction than we now have, since the games are portable, light-weight, robust, and inexpensive, all things that the old apparatus tests were not (Kennedy, Bittner, Harbeson, & Jones, 1983). An encouraging bit of evidence (Kennedy, Bittner, & Jones, 1981) in this connection is an unattenuated correlation close to unity between stabilized levels of practice on Air Combat Maneuvering (ACM) and conventional compensatory tracking, an apparatus test of the older sort. Unfortunately the sample was small. Thomas and Weltman (1981) have designed a part-task video trainer for tank gunnery training. Evaluation of this trainer is currently underway. The Naval Biodynamics Laboratory has studied video games extensively, but to date not in direct relation to job performance (Jones, 1981; Jones, Kennedy, & Bittner, 1981).

The first video games ("Pong," manufactured by Atari, Inc., and "Odyssey," manufactured by Magnavox) appeared on the electronic games market in 1972. A few years later Fairchild Camera and Instruments introduced the programmable video game for home use. Since then, television-computer games for home use have proliferated on all sides. In the last few years a software, video-game industry has grown up in association with the Apple II computer. Literally dozens of games have been programmed for this computer and are available either directly from the originator, through catalogues, or in shops specializing in computer equipment (for example, Radio Shack). Many of the games, for example, "Mountain Pilot" by Clifford Schafer, have primarily to do with cognitive, decision-making skills; they are not fast-action psychomotor tasks as are most of the widely available commercial games. "Mountain Pilot," for example, requires a person to take off, fly a plane over a mountain, and return. The plane must reach altitude before crossing over the mountain, and it will stall if speed and rate of climb are not handled properly. The player must keep track of fuel consumption and if he or she takes too long to reach altitude, there will not be enough fuel to cross the mountain, land, and return to base camp.

Simutis, Baker, Bersh, and Alderman (1979) studied transfer of training in three treatment groups, including two which practiced video games, to the reading of logic circuit diagrams. One group, called a "logic control" group received one hour of instruction on the interpretation and meaning of logical functions and other features of logic circuit diagrams, followed by two hours of practice reading the diagrams. The experimental or "logic game" group practiced the video game Mastermind, presented on the University of Illinois Plato IV system, for one hour, followed by two hours of practice reading circuit diagrams; the latter practice was of the same sort but at a lower level of complexity than the practice afforded the logic control group. The third or "control game" group played a video-game version of Blackjack for an hour, also on the Plato IV system, followed by two hours of practice of the same sort as that received by the second (experimental) group. On the transfer task (reading logic circuit diagrams) the logic game condition was generally intermediate between the logic control group and the third (control-game) group. The superiority of the logic to the control game was not a matter of playing games in general since the two game conditions differed only in the kind of game that was practiced. A second investigation involving troubleshooting of a simulated computer circuit, used essentially the same treatment conditions (Baker, 1981).

Lynch (1981 a & b) has reported extensive work using Atari games in the treatment of brain-damaged patients. The objective in these studies is to facilitate the recovery of perceptual and cognitive function in persons who have experienced traumatic injury to the central nervous system. Lynch's results indicate that video games are helpful in this process but that the results are preliminary.

Recovery of function following traumatic injury to the brain or spinal column is not what one ordinarily understands by "job performance." Nevertheless, the functions at issue are the same as those used in many jobs, and a positive showing for the games in cognitive retraining would augur well for a similar showing relative to conventional jobs.

Job Sample

Eaton, Johnson, and Black (1980) engaged in a three-phase study to investigate job samples as predictors of tank gunnery performance. Three job samples, tracking, sensing, and round adjustment, were chosen for evaluation. In Phase I one of the tracking tasks was found to be related to gunnery performance ($r = -.41$, $p < .05$) and first round hits ($r = -.50$, $p < .01$). Correlations between the sensing task and gunnery approached significance. No significant relationship was found between the round adjustment task and gunnery. These results were generally confirmed by Phase II of the experiment. Correlations between the diamond tracking task and the criteria were between -.41 and -.49 (all significant). Again, correlations between the sensing task and gunnery scores generally approached significance ($r = -.36$, $p < .10$) and one, between sensing errors and gunnery scores was significant ($r = -.41$, $p < .05$). Combining tracking and sensing (signs reversed to produce positive scale) resulted in $r = .64$, $p < .01$. In Phase III a different population of subjects was used, soldiers who had completed eight weeks of Basic Armor Training and soldiers with no training. Again significant relationships were found between the tracking task and gunnery, however, none were found for the sensing task nor for a new task that was introduced, center-of-mass (traversing the turret, laying the main gun quickly and accurately).

The Criterion Problem

It is axiomatic that validity of the criterion and of the predictors is a critical element in the successful undertaking of selection/personnel testing. In line with this, it is well known that the limit of the validity coefficient of a criterion (or a predictor) is the reliability of that measure.

There are several types of validity commonly reported in textbooks of psychometrics. These include: (a) predictive, or the correlation between operator performance on a task (or tasks) and future criterion performance or status; (b) concurrent (diagnostic), or the correlation between test score and a diagnostic criterion status obtained at approximately the same time; (c) content, or the extent to which a task or task battery covers a representative sample of the behavior domains to be measured; (d) construct, or the extent to which a test may be said to measure a "theoretical construct" or trait where theoretical construct or trait is established by convergence of information from a variety of sources; (e) factor, or the extent to which factor analysis indicates tasks as identifying or correlating with a factor; and (f) face, or the extent to which test "looks valid" to subjects who take it, experimenters, or other observers.

In addition to the commonly listed validities above, prospective tests should be evaluated for statistical features, or the extent to which the test has sufficiently good statistical properties that it is useful for prediction. Specifically, high reliability (task definition), and good stability (Bittner, Carter, Kennedy, Harbeson, & Krause, 1983). Test fairness and other standards of testing (American Psychological Association, 1974) were considered during the test candidate selection process. The intrinsic appeal of a test (task validity), although not ordinarily part of a list of validities, can be considered another form. Task validity is somewhat related to face validity (in the sense that job samples are). Tests with high task validity are better measures of a person's ability because he or she is willing to exert more effort. Often in testing, uncontrolled subject variables (fatigue, motivation, past experience) will modify performance scores, because all the subjects do not try to the same extent. One way of establishing task validity that has been proposed elsewhere (Jones, Kennedy, & Bittner, 1981) is through the use of video games. The above validities were considered for all potential test-battery candidates. Ideally, it might be best if all tests possessed all validities. However, in the present investigation, it was necessary to perform trade-offs as choices were made.

Traditional criterion measures that have been used in past research reflect only part of the job. Criteria previously employed have not generally been concerned with decision-making nor have they been studied within the context of tank commander selection.

The research by Eaton and Whalen (1980) was an effort to investigate the criterion problem. Five sensing procedures were used: TCs with rangefinders, trainees with binoculars, researchers with binoculars, trainees with periscopes, and trainees with binoculars on scoring stand. None of the procedures was found to be especially accurate: the two best were only 87 and 86 percent correct. This discrepancy of accuracy in the criterion could lead to discouraging results of a prediction scheme, to say nothing of the validity problem. A video-tape of the firing might be one way of producing reliable measures.

One way the Army assesses organizational performance during field exercises is with the Multiple Integrated Laser Evaluation System (MILES). Vehicles and personnel are equipped with "eyesafe" laser weapons and sensors. Whenever a laser burst is fired and hits an appropriate sensor, the recipient of the burst is placed out of action (destroyed) by virtue of having his own weapon deactivated. Sensors can only be deactivated by appropriate laser stimuli, e.g., a large vehicle such as a tank can only be destroyed by a large weapon and not by small arms fire, while personnel are susceptible to both.

Performance during MILES exercises, held at the end of each BNCOC class, was observed, as part of the present research, to assess feasibility for use as criterion measures. After observing an end-of-course MILES exercise it was apparent that the students were still learning how to operate the system and not exhibiting the level of decision-making behavior required to qualify as criterion measures. In other words, at the end of TC training the students were still concerned with the mechanics of how to smoothly operate the tank and associated equipment, rather than dealing with the importance of terrain utilization and the development of good tactics and strategies. It was, therefore, concluded that MILES exercises per se, conducted upon completion of BNCOC training, could not be used in the present research for performance assessment.

Since direct measurement of field performance of TCs during MILES exercises conducted during the last week of training was found to be unsuitable, other criteria were considered. The notion of using grades as measures of performance was discarded for the same reason as the MILES exercises, i.e., the present investigation was not concerned with performance as a student, but rather with performance following training. Course grades were also deemed inappropriate, for purposes of this investigation, because they tend to be equipment and organizationally oriented (firing accuracy, military bearing, physical fitness, aggressiveness, etc.) rather than directed toward individual performance of concrete job functions in the field.

Over the past two decades significant advances have been made in the area of performance appraisal systems. Smith and Kendall (1963) were pioneers in the development of techniques known as behaviorally anchored rating scales. This concept involves a procedure for construction of evaluative rating scales anchored by examples of expected behavior. The procedure requires appropriate organizational personnel to consider in detail the components of performance for the job in question and to define, in specific behavioral terms, anchors for the performance continua. Smith and

Kendall contend that traditionally psychologists have imposed their own values, interpretations and beliefs about behavior upon the raters, rather than having the raters determine the content of the rating scale. Smith and Kendall maintain that consensus among raters is a prerequisite to the development of a valid rating scale. Raters cannot be expected to use scales with any conviction or agreement, unless they participate in the development of the scales. Raters must indicate, in their own terms, what kind of behavior represents each level of discriminably different characteristics, and which trait is illustrated by each kind of behavior. Furthermore, rating scales must have face validity and the rater must be convinced of the importance of completing the ratings honestly and carefully.

The procedure used by Smith and Kendall was similar to that of foreign language translation, where the translation of one language to another is followed by retranslation by an independent translator. The translation is then examined to determine adherence to connotations as well as to denotations of the original. Smith and Kendall had raters develop and examine examples and classify them as indicative of a given dimension of job performance. Independent judges were then asked to indicate which dimension was illustrated by each example. Strong agreement was reported for a number of examples and scale reliabilities were high (.97).

The behaviorally anchored rating scale approach has been used in a number of academic, industrial, and military settings. Smith and Kendall found it to be promising not only for nursing personnel in hospitals, but for other complex tasks. Campbell, Dunnette, Arvey, and Hellervik (1973) made a distinction between criterion measures that assess individual performance in terms of concrete job functions and those that reflect organizational outcomes several steps removed from actual behavior. Campbell, et al. (1973) were concerned with the assessment of individual performance in terms of concrete job functions and developed a behaviorally anchored rating scale. It was found, based upon a large population of department store managers, that this variant of the scaled expectations procedure produced performance ratings that were not subject to many of the errors commonly associated with such ratings. One of the conclusions was that, potentially, the outputs of this procedure could serve as criteria against which to evaluate predictors for selection and promotion.

A behaviorally anchored rating scale approach was developed by Harari and Zedeck (1973) for the evaluation of faculty teaching. The results support the notion that students are competent and mature raters who separate good teaching from showmanship and popularity and that student opinion, when systematically assessed, can be a key input for evaluation of faculty teaching.

In an attempt to improve the Marine Corps performance evaluation system, Murphy (1980) conducted an extensive review of the behaviorally anchored rating scale literature and concluded that the advantages of the behaviorally anchored rating scale approach over traditional rating formats are: major job components are readily identified; language is clear and unambiguous; pertinent behavior is pin-pointed; agreement is high between rater and ratee; and improvement in performance when used for counseling purposes.

The daily interactions of students in the classroom and in the field result in their getting to know the characteristics and capabilities of each other both socially and as tank crewmembers. These classroom and field interactions provide the opportunity for the establishment of reliable and valid assessment measures, known as peer ratings. Peer ratings are the evaluation of members of a group by one or more other members of the group. Peer ratings have been shown to be good predictors of success or failure in several areas of endeavor, even though the ratings were made by relatively unsophisticated and untrained observers.

Peer ratings have a long history as effective performance appraisal techniques in the military. In many instances peer ratings have been found to be superior to other methods of predicting and assessing performance. In a study designed to investigate peer nominations as predictors of Officer Candidate School (OCS) grades and combat leadership, Williams and Leavitt (1947) reported that the peer ratings were superior to supervisory ratings, and several objective tests, as predictors of both grades and leadership. One reason that the authors attribute to the superiority of the peer ratings over supervisory ratings, is the fact that the students have more time to observe each other than do the supervisors. The other possible explanation given was that the students "know each other in a realistic social context, and they react directly to each other's social dominance behavior". Baier (1947) reported that peer ratings at West Point were more indicative of future officer performance than was scholastic standing.

In a study of Naval OCS graduates from Newport, it was found that none of the selection variables were predictive of later success on-the-job, as measured by officer fitness reports. However, a significant correlation of .23 was obtained between peer ratings administered in the 11th week of OCS and fitness reports of officers assigned to both shore and fleet billets (La Gaipa, 1961).

Peer ratings were administered, by Rigby and Ossorio (1959), early in the training of Navy and Marine Corps female recruits, and it was found that the peer ratings discriminated between successful and non-successful trainees.

Doll (1963), in an attempt to predict failure to complete flight training, found that peer ratings by cadets who live together "around the clock" were more effective predictors than ratings by officers who only attended class together. Jones and Doll (1974), however, found peer ratings of pilots and Naval flight officers to be effective as criterion measures for initial acquisition of targets. In this case the groups were not composed of students who closely associated daily, but aircrew members who occasionally flew together.

In order to predict performance, statistical reliability of the predictors and the criteria must be established. Assuming absolutely reliable predictors, the outcome of a prediction scheme can only be as good as the reliability of the criteria. Pass/fail is frequently used as a criterion of success in performance prediction programs, however, the effectiveness of this approach suffers when the pass/fail ratio is not somewhere near equal. Since the percentage of failures in a TC course is low, passing or failing the course would not be a suitable criterion. However, passing or failing a specific task, is considered worthy of investigation as a criterion.

Task Analysis Preliminary to Prediction

Assessing leadership is difficult because criterion measures are elusive. Sometimes the opinion of "experts" is relied upon. In the military this is frequently an evaluation of subordinates by superiors. One approach that has achieved a modicum of success, developed by Flanagan (1954), is the critical incident technique. Klemp, Munger, and Spencer (1977), using critical incident interviews refined for structured "behavioral event" analysis, evaluated: (a) Task Achievement (concern for achievement, takes initiative, sets goals, coaches, technical problem solving); (b) Skillful Use of Influence (concern for influence, influences, conceptualizes, team builds, rewards, self-control); (c) Management Control (plans and organizes, directs, delegates, optimizes, monitors results, resolves conflicts, gives feedback); (d) Advising and Counseling (listens, understands, helps, positive expectations); and (e) Coercion (coerces, negative expectations, disciplines, acts impulsively, fails to resolve conflicts), for the analysis of leadership and management competencies of commissioned and noncommissioned naval officers. They found that four factors differentiate between superior and average leadership and management performance. These were: Task Achievement, Skillful Use of Influence, Management Control, and Advising and Counseling. Coercion did not distinguish average from superior performers. The critical incident technique, while providing certain necessary information for use in the development of predictors of TC performance, does not provide the specific detailed information required, therefore, a special form of task analysis was developed for the present study.

Task description and analysis are rarely, if ever, done for their own sakes; they serve a larger purpose or purposes. One of these larger purposes is training, others are organization of the work process and equipment design. Prediction is still another reason for doing a task analysis; but it differs radically from the three earlier mentioned reasons. That is, a task analysis preliminary to prediction is categorically different from a task analysis carried out with a view to training, organization of work, or equipment design. It not only serves another end but that end dictates a different kind of analysis than is usually performed, with different requirements and different criteria for success.

The gist of this difference is that in task analysis preliminary to prediction no job element is relevant unless some workers <u>fail</u> to perform it adequately. Consider driving a car and suppose that everyone without exception knows how to start a car under normal circumstances. In such a case starting the car cannot possibly be of any help in predicting who will drive well and who badly because nothing can predict a difference among individuals except another individual difference, and starting the car, under these assumptions, shows no variation from one individual to another. This point is well understood in psychometric theory. Items are not usually even considered for inclusion in a test unless the difficulty level, that is, the percent failing the item, lies somewhere between 15 and 85 percent. If only five percent of the criterion population fails an item (or if only five percent passes it) there is not enough variation on the job element to make it useful in prediction no matter how badly the rare individual who fails the item may do on the overall task. Starting a car is, no doubt, essential to driving one well. Nevertheless, it has no place in a task analysis preliminary to prediction unless an appreciable proportion of persons fail to do it at least passably well.

When a task analysis is carried out with a view to training, no important part of the task can be safely ignored. In training, for example, everyone must learn to do even the simplest parts of the job and the fact that no one on the job ever fails to perform a given task component is insufficient basis for not teaching it. The same is true when task analyses are carried out preliminary to the organization of work or equipment design. Perhaps a particular, no doubt, elementary part of a job is rarely, if ever, performed unsatisfactorily. It still requires a certain amount of time to be performed and equipment must still be designed so that it can be performed safely and efficiently.

No variety of task analysis to date has taken this difference into account (McCormick, 1979; Miller, 1962), not even those varieties that take a differential approach (Fleishman, 1975a & b). The task itself is still analyzed exclusively in terms of important components or, in the differential case, what appear to be important abilities in successful performance of the overall task. In the latter case, if a task clearly involves visual search, then visual search as an ability is said to be required and so, in some sense, it may be; but if no one fails adequately to carry out the visual search operations required in the task, then visual search as an ability will be entirely useless in predicting who does well and who poorly on the overall task.

For these reasons, because the purpose of this research was ultimately to predict performance as a TC and no existing variety of task analysis serves this purpose, it was necessary to first determine what a proper task analysis should be. Once that was done, applying the conclusions to TC performance as a particular case was straightforward.

Statement of the Problem

Most military tasks and parts of tasks are treated on a pass/fail, satisfactory/unsatisfactory basis; the mission is either accomplished or it is not accomplished and a particular part of that mission is either performed as doctrine and training manuals say it should be performed or it is not. Hence, a dichotomous format was adopted in stating the problem. The dimensions of that problem have already been indicated. On the one hand, there is the mission; on the other, a part of that mission or job element. The study units are workers or, in this case, tank commanders. The problem, therefore, takes the form of a two-by-two table:

		Mission		
		Success	Failure	
Job	Success	A	B	P=A+B
Element	Failure	C	D	Q=C+D
		R=A+C	S=B+D	N=A+B+C+D

where, A, B, C and D refer to the number of persons in a series of appropriate observations who perform as indicated. In this format the problem of task analysis has two dimensions. The first corresponds to item difficulty in psychometric theory and is given by Q/N, the proportion of persons who fail the job element. This dimension is "relevance," in the sense that a job element is relevant only if Q/N is appreciably greater than zero and less than unity. The second dimension is the "importance" of the job element to mission success. This dimension has two concrete forms: the probability of mission failure if one succeeds on the job element

$$P\ (MF \mid ES) = q_s = \frac{B}{A + B} \quad , \tag{1}$$

and the probability of mission failure if one fails on the job element

$$P\ (MF \mid EF) = q_f = \frac{D}{C + D} \quad . \tag{2}$$

Importance is the difference between these two probabilities, that is, $(q_f - q_s)$.

The problem now is how to put these two dimensions together into a single statement of the functional relationship between success/failure on the job element and success/failure on the mission. How does one model this relationship or, better, how does one model it best for the purposes of task analysis?

Attributable Risk of Mission Failure

The functional relation between two dichotomous variations can and has been modeled in many ways. A psychologist thinks first of the tectrachoric correlation coefficient, fourfold point correlation, χ^2, or the contingency coefficient. None of these models, however, will serve our purpose, because the two dimensions of the problem either do not appear clearly as such or do not relate to one another in an intuitively clear and intelligible form. In task analysis one cannot expect always to have a complete data base, usually one does not. In such cases, and the present is a case in point, one must rely on the opinion of subject matter experts (SMEs) as to how relevant or important a job element is. That being true, the model adopted must (a) involve relevance and importance, (b) involve them in an intuitively simple way, and (c) involve no other consideration that varies from one risk factor to another. In this report, a detailed analysis of the familiar measures of association will not be presented. Suffice it to say that none of them meet the requirements of this study. To find a model that does one must look elsewhere. When one does, it is found in epidemiology! It is called "attributable risk" and was developed to formalize how much of lung cancer is attributable to smoking or, more generally, how much of a given disease is attributable to any given risk factor. First the model as it was developed in epidemiology will be explained and then translated to the model for task analysis.

In its original epidemiological context, whether or not a person has a disease corresponds to mission failure/success. Similarly, whether or not a person was exposed to a given risk factor corresponds to failure/success on the job element. The reasoning then proceeds as follows. The probability of developing the disease if one has not been exposed to the risk factor is q_s and the probability of developing the disease if one has been exposed to the risk factor is q_f. Therefore, the risk of developing the disease attributable to the risk factor, that is, the <u>additional</u> risk if one has been exposed to the risk factor is $(q_f - q_s)$. This risk applies, of course, only to those persons who are exposed to the disease, that is, Q. Hence, the number of cases attributable to the risk factor is $(q_f - q_s)\ Q$. Altogether, however, one has to account for S cases of the disease. Hence, the proportion of cases attributable to the risk factor is

$$\frac{(q_f - q_s)\ Q}{S} \qquad (3)$$

On the basis of his own data, Levin (1953) concluded that almost 80 percent of all cases of lung cancer was attributable to smoking.

The translation of this reasoning to the proposed analysis is uncomplicated. To begin with, one must know how much the probability of mission failure is increased by failing a given job element. The answer is $(q_f - q_s)$. This risk applies only to those workers who fail the job element — and this, of course, is where relevance enters the picture. Thus, the number of mission failures attributable to failure on the job element is

$$(q_f - q_s)\ Q \qquad (4)$$

and the proportion of mission failures attributable to failure on the job is formula (3).

Since it is more intuitive to deal with proportions than absolute numbers, attributable risk can be rewritten as

$$AR = (^{Q}/_{N})\ (q_f - q_s)\ \ /\ \ (^{S}/_{N})\ . \qquad (5)$$

The first expression in the numerator is relevance. The second is importance; it tells "how much difference it makes" if a person fails or succeeds on the job element. The expression in the denominator is the proportion of persons who fail on the mission and does not change from one job element to another. It is a constant.

Tank Commander Performance Prediction

How this model is used in practice depends primarily on the data base available. Three situations may be distinguished. In the first the data base is complete. A representative sample of workers has been studied and information is available in each case as to success/failure on the job element and mission outcome. In the second situation information is available as to mission outcome; one knows how frequently mission failure occurs, but not as to the job elements. In the third situation no information is available other than SME opinion, not even as to the frequency of mission failure. In this third situation, and the present investigation is a case in point, each SME is asked to rate both importance ($q_f - q_s$) and relevance (Q/N). Relative AR is then obtained by multiplying the two ratings together.

In all three cases a list of job elements (tasks or subtasks) was needed before the analysis can be conducted. Fortunately, it was not necessary to construct a list of tank commander tasks. That step had already been carried out for the TC job by Drucker & O'Brien (1981, pp. 184-185). The Drucker & O'Brien list was especially appropriate in that it concentrated on leadership as contrasted with equipment-related tasks. The Drucker & O'Brien list, which contains 71 tasks, was reduced to 32 tasks by having SMEs eliminate items not relevant to this analysis and consolidating others. The reduced list is presented in Table 1.

The task ratings were carried out in two sessions. In each session two interviewers interviewed a platoon leader and his platoon sergeant. No attempt was made to separate the judgments of the two tankers. When a task was discussed, both men expressed their views and the two interviewers separately noted what he or she took to be their consensus.

Table 1
Thirty-Two Tank Commander Leadership Tasks

Task Number	Task Description
1	Acquire targets
2	Analyze and utilize terrain
3	Announce feeding and rest plans
4	Assign sectors of observation
5	Clarify mission
6	Identify and select tank targets
7	Ensure maintenance
8	Post and ground guard
9	Select and prepare alternate position
10	Prepare and complete battle position
11	Direct cease fire
12	Ensure communications check
13	Maintain correct interval with other tanks
14	Direct main gun be oriented
15	Carry out movements in formation correctly
16	Ensure readiness actions are conducted
17	Direct rapid movement into assigned area
18	Direct tank be camouflaged
19	Engage surprise targets
20	Engage targets in assigned sector
21	Identify withdrawal route
22	Initiate range card preparation
23	Request, monitor, and adjust indirect fires
24	Prepare for 3-man crew operations
25	Provide supply and maintenance status to platoon leader
26	Report personnel status
27	Select good fields of fire
28	Shift fire on order
29	Submit SITREP
30	Submit SPOTREP
31	Tie in tank with elements on left and right
32	Wait for order to open fire

The ratings themselves were of two sorts. The first had to do with how frequently the task was performed inadequately. The two tankers were simply asked, "How frequently in your experience is task X performed inadequately?" The two interviewers listened to the discussion that followed and then made their ratings. Five standard points were provided: .01, .04, .09, .17, and .33. For each task each interviewer rated the two tankers' judgment at one of these five points. This set of points was used rather than exact values of $.02 \times 2^n$ (.02, .04, .08, .16, .32) because it was desired that the platoon leaders and sergeants think of the rating points as absolute proportions of indefinite performances rather than a rating scale and it seemed more likely that they would do so if the rating points did not conform exactly to $.02 \times 2^n$.

The second rating concerned the importance of the task to overall mission success or failure. The two tankers were asked, "How important is task X to overall mission success or failure?" For this rating four standard points were provided:

- Virtually rules out mission success, .80
- Seriously jeopardizes mission success, .60
- Complicates mission success, .40
- Mildly impedes mission success, .20.

Occasionally, the interviewers would indicate other rating points, for example, .70, if they felt that the judgement of the two tankers lay between standard points. More frequently, the interviewers would ask the tankers to compare a task under discussion with one they had already discussed. This latter tactic was especially helpful in getting at relative importance, since tankers generally regard every task they are assigned as of utmost importance. When pressed, however, they readily concede that some tasks are more important than others.

The results of the ratings appear in Table 2. For each interviewer in each session three columns of figures are presented. The first column contains that interviewer's rating as to frequency of inadequate performance of the task; the second column contains that interviewer's rating as to importance of the task to mission success or failure; the third column contains the risk of mission failure attributable to failure on that task. This last column is calculated by simply multiplying the first two columns together.

Table 2

Frequency of Inadequate Performance (FREQ.) Importance to Mission Success or Failure (IMP.) and Attributable Risk of Mission Failure (AR) for each of 32 Tasks, By Session (May 20 or 21) and Interviewer (Jones or Black)

	MAY 20						MAY 21						AVERAGE AR		
	JONES			BLACK			JONES			BLACK			5/20	5/21	Both
TASKS	FREQ	IMP	AR	FREQ	IMP	AR	FREQ	IMP	AR	FREQ	IMP	AR	AR	AR	AR
1	.01	.70	.007	.01	.80	.008	.01	.80	.008	.09	.80	.072	.008	.040	.024
2	.33	.80	.264	.33	.80	.264	.17	.80	.136	.04	.80	.032	.264	.084	.174
3	.01	.20	.002	.01	.40	.004	.01	.40	.004	.01	.60	.006	.003	.005	.004
4	.17	.70	.119	.04	.80	.032	.01	.60	.006	.01	.50	.005	.076	.006	.041
5	.17	.60	.102	.33	.70	.231	.04	.40	.016	.01	.60	.006	.166	.011	.088
6	.33	.60	.198	.17	.80	.136	.01	.60	.006	.04	.60	.024	.167	.015	.091
7	.17	.40	.068	.17	.60	.102	.33	.80	.264	.33	.80	.264	.085	.264	.174
8	.01	.20	.002	.01	.20	.002	.01	.60	.006	.04	.70	.028	.002	.017	.010
9	.33	.40	.132	.33	.60	.198	.17	.60	.102	.33	.80	.264	.165	.181	.173
10	.33	.20	.066	.17	.60	.102	.01	.80	.008	.04	.80	.032	.084	.020	.052
11	.01	.20	.002	.01	.20	.002	.01	.20	.002	.01	.40	.004	.002	.003	.002
12	.01	.80	.008	.01	.80	.008	.04	.60	.024	.01	.70	.007	.008	.016	.012
13	.09	.20	.018	.04	.20	.008	.09	.40	.036	.17	.40	.068	.013	.052	.032
14	.01	.20	.002	.01	.20	.002	.01	.40	.004	.01	.40	.004	.002	.004	.003
15	.17	.40	.068	.17	.70	.119	.09	.60	.054	.17	.80	.136	.094	.095	.094
16	.04	.60	.024	.04	.40	.016	.17	.60	.102	.33	.80	.164	.020	.181	.100
17	.01	.40	.004	.01	.20	.002	.01	.60	.006	.04	.60	.024	.003	.015	.009
18	.01	.20	.002	.01	.40	.004	.04	.60	.024	.01	.70	.007	.003	.016	.010
19	.04	.60	.024	.01	.60	.006	.01	.80	.008	.04	.85	.034	.015	.021	.018
20	.01	.40	.004	.04	.60	.024	.01	.60	.006	.09	.80	.072	.014	.039	.026
21	.33	.60	.198	.33	.80	.264	.17	.80	.136	.33	.80	.264	.231	.200	.216
22	.01	.60	.006	.01	.60	.006	.17	.80	.136	.17	.80	.136	.006	.136	.071
23	.04	.40	.016	.01	.20	.002	.17	.80	.136	.33	.80	.264	.009	.200	.104
24	.01	.20	.002	.01	.20	.002	.04	.40	.016	.17	.60	.102	.002	.059	.030
25	.01	.60	.006	.01	.80	.008	.33	.80	.264	.33	.80	.264	.007	.264	.136
26	.01	.40	.004	.01	.80	.008	.01	.40	.004	.01	.60	.006	.006	.005	.006
27	.04	.80	.032	.04	.80	.032	.09	.40	.036	.17	.80	.136	.032	.086	.059
28	.01	.40	.004	.01	.80	.008	.01	.60	.006	.01	.80	.008	.006	.007	.006
29	.17	.40	.068	.09	.40	.036	.04	.60	.024	.01	.60	.006	.052	.015	.034
30	.01	.40	.004	.01	.60	.006	.04	.60	.024	.01	.60	.006	.005	.015	.010
31	.01	.20	.002	.01	.20	.002	.09	.40	.036	.17	.60	.102	.002	.069	.036
32	.01	.20	.002	.01	.20	.002	.01	.80	.008	.04	.80	.032	.002	.020	.011

The correlation between attributable risk (AR) for the 32 tasks as assessed by the two interviewers was .89 within the first session and .80 within the second. These correlations indicate good inter-rater reliabilities.

The next to last columns on the right and the one just before that present AR averaged over the two interviewers for each session. The correlation of the ARs between the two sessions is .25, indicating that the two pairs of tankers had definitely different views as to the frequency of inadequate performance and importance to mission success or failure of the 32 tasks. The last column in Table 2 contains the average AR for the two sessions. The low correlation between average AR in the two sessions does not necessarily invalidate the overall average AR as an indicator of criticality. It merely means that this indicator represents an average of divergent views, not a consensus.

Table 3 presents the overall average AR for the 10 most critical tasks. With respect to tasks 21, 9, and 15 the teams of tankers of the two sessions were in essential agreement. With respect to tasks 2 and 7 there was considerable but not extensive disagreement; both teams judged the two tasks to be more critical than the average task but not to the same extent. With respect to tasks 25, 16, 6, and 5 the two teams disagreed widely, their ratings differing in direction from the mean as well as extent. Task 23 presented a special problem. The two teams disagreed widely on this task but, at the same time, were agreed that tank commanders were not trained to request, monitor, and adjust indirect fires. The difference was that one team regarded the lack of training as an "extenuating circumstance" and the other did not.

The four most critical tasks in Table 3 are alike in that all four have to do with "thinking ahead." Tasks 21, 2 and 9 are connected; they all have to do with what happens after the tank has assumed its primary battle position and fired its main gun. At that point the tank must move to an alternate battle position; and it is here that the main differences between good and bad tank commanders (TCs) show up. A good TC has selected and prepared his alternate battle position in advance. He has also identified the route he will follow in withdrawing from his primary battle position. Finally, he has chosen his route so as to take maximal advantage of the terrain; specifically, he does not expose his tank to enemy fire in the process of moving to an alternate battle position. A poor TC fails to think beyond the immediate situation in which he finds himself; all steps after the first are improvised.

Table 3
The Ten Most Critical Tasks as Judged By Two Teams of Tankers and Evaluated By Two Interviewers

Tank Number	Description	AR
21	Identify withdrawal route	.216
2	Analyze and utilize terrain[1]	.174
7	Insure maintenance[1]	.174
9	Select and prepare alternate position	.173
25	Provide supply and maintenance status[2]	.136
23	Request, monitor, and adjust indirect fires[3]	.104
16	Insure readiness actions[2]	.100
15	Carry out movements in formation	.094
6	Identify and select tank targets[2]	.091
5	Clarify mission[2]	.088

[1] There was considerable disagreement between May 20 and May 21 on this task.

[2] There was widespread disagreement between May 20 and May 21 on this task.

[3] TCs cannot do this because they are not trained to do it.

Task 7, ensure maintenance, includes at-halt and after-operations maintenance as well as emergency repairs. This task also involves thinking ahead. A good TC maintains his tank at all times; he does not wait until something malfunctions before taking action.

This general theme of thinking ahead is echoed in at least two of the remaining six tasks in Table 3: task 25, provides supply and maintenance status, and task 16, ensure readiness actions. Task 16 refers mainly to pre-combat checks. Even task 5, clarify mission, involves a component of thinking ahead.

The TC Video Test

In view of the evidence that the job sample approach has merit, and that paper-and-pencil tests have not been found to be reliable predictors of TC performance, and since grades could not be used as predictors in the present investigation because the purpose was to select TCs prior to training, a totally different method was developed for using job samples as predictors.

Cognitive decision-making skills or leadership, motivation, and training effectiveness are difficult to evaluate for several reasons: (a) they are difficult to observe; (b) outcome indicators generally reflect other factors as well, for example, equipment-oriented skills, crewmember effectiveness, etc.; (c) the exercises necessary to evaluate TC performance may be lengthy and expensive; and (d) there may be no feasible way to eliminate practice effects.

Since the last point reflects a recent (though currently acute) concern in psychological tests and measurements, it bears further elaboration. Educational tests do not ordinarily provide feedback as to knowledge of results. A person taking a typical IQ or achievement test does not know whether the answers that he or she makes are correct. Consequently, practice effects, while they exist in educational tests, are not large (Messick & Jungeblut, 1981). Performance tests, however, are another matter because here knowledge of results is frequently unavoidable. In a hidden figures test, for example, it is impossible not to know when one has recognized a face in the picture. In other tasks (and the TC's is one of them), knowledge of results is partial. A TC knows a good deal about how he is doing but not everything. The effect, in all probability, will be marked improvement simply as a consequence of performing the exercises necessary to evaluation. If the same TC were to carry out the same exercises another time, he would probably do better.

In a task that is to be used as a predictor, sizeable practice effects present serious problems. How a person performs on such a task may have little to do with how he or she would perform given extended practice (Jones, 1980, 1972). The "test" amounts to a trial early in practice, when the learning curve is rising most steeply and performance is most vulnerable to practice effects. How well a person does on such a test may have primarily to do with previous exposure to the same or similar tasks and relatively little to do with that person's abilities. By the same token, how a person performs on a second test (after a month or more) will depend very much on what happens to him or her in the interim, how much relevant exposure he or she has in the meantime. The upshot is poor retest reliability and, with it, poor predictive validity. If a test is to be useful predictively, performance on it must characterize the subject (not his or her exposure history) and, where practice effects are large, this requirement may well not be met.

Unfortunately, the remedies for practice effects are limited. One can select tests that are already overpracticed or one can continue practice until a subject's performance is improving at a slow and regular rate or not at all. Neither of these possibilities is readily accessible in job sample situations. For these reasons it was decided to sample some of the behavioral job-elements judged to be important, and not to give the subject feedback on the actual measures being recorded.

Although there has been a very recent proliferation of video games, including tank games, it was decided not to use any existing games (with two exceptions, noted below) as predictors, but rather to commission a highly skilled video-task programmer, Mr. Shafer, to create a game to our specifications. Several points should be underscored.

First, the video task is not a "game" in the sense of a competition. It requires a subject to complete a mission, much like "Mountain Pilot." The player does not win or lose but accomplishes or does not accomplish the mission tasks assigned him.

Second, performance is assessed along several dimensions. There is not just one overall indicator of performance. A subject is assessed on several of the functions he has been asked to carry out. A TC, for example, receives one score for "staking," i.e., placing a stake at a position to which he might later want to return, and a very different score for time spent in line-of-sight to the target (unobstructed view between own tank and target).

Third, the video task is intended to complement other predictors, not substitute for them. The video task focuses on cognitive, decision-making job elements. It was not planned to assess leadership, motivation, or training effectiveness.

Synopsis of the Research Plan

Two experiments were designed to investigate the effectiveness of a test battery for predicting TC performance. One of the tests was designed as a form of job sample measure. The job sample approach assumes that the applicant is acquainted, to a significant degree, with the job under consideration, therefore, in the present investigation familiarity with the duties of the TC was a prerequisite in the selection of a population for testing. In order to test enough students to produce statistically meaningful results related to performance on the job sample test, it was necessary to test two different populations. The two populations selected for testing were very different in that members of the first group were young officers who were well educated but relatively inexperienced as tank crewmembers, while the second group of enlisted sergeants were less well educated but were older and experienced as tank crewmembers.

Subjects

Experiment 1 was planned because of an inadequate number of students available for testing in Experiment 2. Students selected to participate in Experiment 1 were young officers (lieutenants) who had just completed a platoon leader course which included tactical training and decisions and planning faced by a TC, as well as other instructions on how to command a tank. During the course the officers received some hands-on experience and the course culminated in a 10-day exercise in the field. These students had not had the extensive experience that is desirable in using the job sample approach to prediction, however, it was believed that they had sufficient experience and training to serve as a representative tanker population, for the evaluation of the effectiveness of the job sample based test described below.

The students chosen to participate in Experiment 2 were enlisted men (E5 and E6) who had been selected for TC training. They were experienced as tank drivers, loaders, and gunners. Although these students had never previously been TCs, nor received formal training to become TCs, it was felt that their background and experience as tank crewmembers qualified them as good candidates for the job sample approach.

Predictor Tests

The predictor tests chosen for this research were five paper-and-pencil tests, two commercially available video games, and a microprocessor job sample based test (called "Tank Commander"). The paper-and-pencil tests were Hidden Patterns, Map Planning, and Paper Folding, from the ETS battery of factor-referenced cognitive tests (Ekstrom, French, Harmon, & Derman, 1976), Mathematics, and Grammatical Reasoning. The two video games were Air Combat Maneuvering (ACM) and Touch-Me.

The three tests from the ETS battery have had extensive theoretical work behind them, and are largely marker tests to help with the interpretation of outcomes from the research. The goal was to select tests which possess factor structure and content which would seem likely to be related to the task content which surfaced during the task analysis. Moreover, the ETS tests have some face validity, and were assumed to have predictive and metric validity. Because of the fashion in which the various theories are accommodated within the Ekstrom battery, construct validity was also available for these tasks.

The Mathematics and Grammatical Reasoning tests were selected mainly because they had excellent metric properties (early stabilization and high task definition). Additionally, but less important, they were seemingly factor pure (Carter, Kennedy, & Bittner, 1981) and independent, therefore, if they were correlated with the criteron, the correlations would be informative and the outcomes would be interpretable. These two tests have also been predictive over the years for various occupations, and changes in some environments, and so they have a sensitivity to treatments. Mathematics and Grammatical Reasoning have not been shown to be related to performance on anything like commanding a tank. Content validity may be present in Math, but not for Grammatical Reasoning. In the Carter et al. (1981) study, Grammatical Reasoning was shown to correlate, but negatively, with Spatial Visualization from the Ekstrom et al. (1976) ETS battery. Spatial Visualization, a test with some apparent relationships to commanding a tank, was considered, but in the Carter et al. (1981) research Spatial Visualization did not stabilize quickly and had low reliability.

The Atari Combat ACM task has been demonstrated to have predictive validity for Navy carrier landing performance (Lintern & Kennedy, 1982), and its metric properties are well known and excellent (Jones, Kennedy, & Bittner, 1981). Because ACM is related to more traditional measures of manual control (Kennedy, Bittner, & Jones, 1981), it also has factor validity. The obvious content of the task indicates some face validity in the sense that ACM has a military theme. Construct validity can be inferred through the manual control relationships.

Touch-Me, the other video game, possesses factor validity (Underwood, Borach, & Malmi, 1977; Kennedy, Andrews, & Carter, 1981) and concurrent validity for certain types of diagnostic ills. Its metric properties are good (Kennedy, Andrews, & Carter, 1981) and if short-term memory is an inherent characteristic of commanding a tank, one might say Touch-Me has factor and perhaps construct and face validity.

The Tank Commander video test that was designed for this experiment was created to serve as a job sample (Guion & Ironson, 1979), and it was possible to introduce job sample/realism. Therefore the Tank Commander test has face validity and to some extent concurrent validity. It was not possible to know in advance about the metric properties of the TC test, since it was only just developed, but attention was paid to measurement. Moreover, factor validity could only be estimated, because of the interest in tapping several underlying factors simultaneously, as shown in the task analysis preliminary to prediction. Because of the heterogeneity of the task, the Tank Commander test could be expected to have content validity, but it is recognized that this sometimes works at odds with other forms of validity.

The Tank Commander test was based upon the job sample selection technique, which requires an applicant to produce examples (samples) of ability to perform a particular task. Several aspects of commanding a tank, observed to be frequently failed during field exercises, were incorporated into the Tank Commander test.

The Tank Commander map display can be thought of as being superimposed on an unseen grid. The grid consists of 60 horizontal squares by 40 vertical squares. The squares were set up to correspond to 10 kilometers. Each square, or block, was programmed to: (a) be blank; (b) be colored (representing terrain features); (c) be in line-of-sight to the target (unobstructed view between tank position and target); (d) be in light-of-sight to a bridge; (e) have a probability (of hitting the target) of 0 to 9; (f) be designated as an edge square (good staking position); or (g) be designated as a hazard square (within range of enemy weapon). Hazard, edge, and probability features are not visible during operation of the simulation. Student performance was assessed by analysis of automatically recorded data. The microprocessor recorded the measures below in data files, by the following titles:

1. LOS. The total number of seconds spent in line-of-sight to the target.

2. LOSX. A cumulative score was derived by calculating the number of seconds spent in each LOS block (of the display grid) multiplied by 10,000 and divided by the distance (range in meters) between each block and the target.

3. LOSP. The cumulative number of seconds spent in any block in LOS multiplied by the probability associated with each block.

4. HAZ. The number of seconds spent in LOS within range of the enemy weapon, i.e., hazardous area, was recorded.

5. STAKES. The value of planned battle positions was determined by evaluating the positions of stakes placed by the student. Stakes were placed to either protect a bridge, or preparatory to attack on the target. The scores were: (a) the number of stakes placed; (b) the XY coordinates of each stake; (c) the time each stake was placed; (d) the hit probability

associated with each stake; and (e) whether or not the stakes were placed on LOS or HAZ squares (poor staking positions), or on EDGE squares (good staking positions).

6. FINAL LOCATION STAKED. An indication of whether or not the final location of the tank was staked when the simulation ended.

7. PROB. The hit probability associated with the positions of each of the first three shots.

8. AVE. The average probability of hit of the first three shots fired.

9. SHOT. The XY coordinates of the tank and the time of each of the first three shots fired.

10. MOVED AFTER SHOT. An indication of whether the tank moved after the second shot was fired.

11. SECONDS MISSED ASSIGNED TIME. The instructions called for the student to fire at a specific time. The difference between the time of the first shot and the assigned time was calculated and recorded.

Hypotheses

Since the purpose of this research was selection, emphasis was placed upon identification of students most likely to be poor TCs and hypotheses were formulated accordingly. The following hypotheses were based upon the results of the task analysis:

1. Poor TCs would receive higher scores than good TCs on line-of-sight measures, i.e., LOS, LOSX, and LOSP.

2. Poor TCs would spend more time in hazardous areas than would good TCs, consequently poor TCs would receive higher HAZ scores.

3. Poor TCs would place fewer stakes and select inferior staking positions. Furthermore, poor TCs would not plan ahead far enough to have prepared a staked position from which the final shot was fired as often as the good TCs, therefore, poor TCs would receive inferior scores for STAKES and fewer positive responses to FINAL LOCATION STAKED.

4. Poor TCs would select inferior firing positions, therefore, would receive lower scores for PROB and AVE, than would good TCs.

5. In order to evaluate the performance of a student after firing but missing the target, the computer was programmed so that the target could not be destroyed until after the first three shots were fired. A good TC never remains at a position after two quick shots (at most) have been fired because his own position has been revealed to the enemy. It was, therefore, hypothesized that the poor TC would more often remain in one position and continue to fire, and receive fewer MOVED AFTER SHOT scores, than would the good TC.

6. The poor TCs would deviate from assigned time to fire more often than would good TCs, therefore, would receive higher SECONDS MISSED ASSIGNED TIME scores.

Criterion Measures

Officer students were ranked from best to worst by the instructor staff following the 10-day field exercise at the end of the course. Experiment 1 was designed to use class standing as the criterion.

The second experiment was designed to measure the performance of the sergeants after completion of training. The purpose of the task analysis described earlier was twofold. First to identify factors (job samples) that TCs frequently fail to perform satisfactorily, for incorporation as predicators into the Tank Commander test. Secondly, the task analysis identified those aspects of MILES exercises that would serve as criterion measures of performance in the field.

It was apparent from discussions with the Basic Non-Commissioned Officer Course (BNCOC) cadre that they knew the capabilities of each of their students. The problem then was to develop a scheme for eliciting the needed information from the cadre in a manner that could be expressed in a quantitative fashion. The typical training evaluation method of asking instructors to rate each student in terms of specific field performance behavior relative to "average" was deemed inappropriate for a number of reasons. First of all, in the assessment of performance on a complex task, such as system management (commanding a tank), it is difficult, if not impossible, to define "average." More importantly, the terrain utilization factors identified by the task analysis can best be described in terms of what the student specifically did, rather than as good or bad end-products. For example, it is necessary to know what lead up to a decision to plan a particular battle position, rather than looking at the planned position and attempting to make a judgment as to the merits of that position. A chess move can best be evaluated by a kibitzer when the strategy of the player is known. In a given tactical situation the task components, or job dimensions, which will lead to the achievement of successful results need to be identified in order to be able to develop a scheme for measurement of those job dimensions. Identification of the job dimensions was accomplished by the task analysis. Rating scales were developed for performance appraisal, as well as peer and supervisory rankings.

Approach

In the first experiment, it would have been inappropriate to administer a job sample test to the young officer students prior to the start of training because they would have had little or no knowledge of what is expected of a TC. This experiment was, therefore, of necessity designed to assess performance on both the predictor test, Tank Commander, and the criterion, after training was completed. The students selected for this experiment did not represent the population for which the experimental selection battery developed for the present research was intended, therefore, Tank Commander was the only predictor test administered to this group.

The second experiment was designed for collection of predictor test data prior to the start of training, and to collect criterion data after completion of the final MILES exercises at the end of training. All eight predictor tests were administered to the students before they started training and criterion data were collected after completion of training.

Experiment 1

Method

Subjects

The performance of 40 male Army officers was evaluated. They were students in the Armor Officer Basic (AOB) course. These students were trained as platoon leaders and TCs. Upon completion of coursework, the students participated in an end-of-course, 10-day field exercise called the "Ten-Day War." Instructors were requested to place the students into one of three categories; high performers, average performers, or low performers. Of the 40 tested, 14 were from the top of the class, 13 from the middle, and 13 from the bottom.

Materials

Testing was conducted in a room large enough to accommodate one student, one test administrator, and the testing equipment.

Predictors. The prototype computer-video test, designated Tank Commander (described above under "Synopsis of the Research Plan" and in more detail in Appendix A), was developed for use as a predictor of TC performance. The game consists of a colored map type display, on the monitor, of various terrain features such as mountains, hills, roads, rivers, etc. The map consists of horizontal and vertical squares set up to correspond to 10 kilometers. Each square, or block, can be programmed to be blank, colored (representing terrain features), in line-of-sight to the target (T LOS), in line-of-sight to the bridge (B LOS), have a probability of 0 to 9, be an EDGE square (good staking position), or be a hazard (HAZ) square (within range of target weapon). Hazard, edge, and probabilities are not visible during operation of the simulation. A "tank" is displayed as the cursor at the center of a white cross (X) and is programmed to approximate a maximum speed of 90 kilometers per hour. The tank is moved around the map by means of a joystick. Three buttons are used to fire, place a stake, or advance the clock to attack time when the student is satisfied with planned battle strategies and is ready to fire. In order to speed up the game, computer time is three times as fast as real time. Measures of performance were concerned with time spent in line-of-sight, in hazardous zones, appropriateness of time of firing and the value of prepared battle positions (stake placement).

Criterion Measures. Class standing at the end of the 10-day war was the criterion used for assessment of the relationship between performance on the Tank Commander test and performance as a student platoon leader/tank commander.

Procedures

Students were provided with a copy of the Privacy Act and filled out a questionnaire pertaining to biographical information and the use of video games (see Appendix B). The test administrator read the instructions to each student (see Appendix C). Three computer maps were used: a Demonstration Map, Map A, and Map B. See Appendix D for examples of computer displays. Initially the demonstration map was displayed on the monitor, the student was trained to operate the equipment, and the meaning of the displayed information was explained. Student interaction was encouraged, e.g., "as you go around the mountain notice the bridge line-of-sight (B LOS) light come on," or, at the appropriate time, "why do you not presently have line-of-sight to the target?", or "why is the READY light not on now?" Total time for each student to play the Tank Commander game, including training and the demonstration trial, was about 45 minutes.

After the student had learned how to operate the game and had completed the practice trial, he was given an operational order (OPORDER) to read, while the test administrator set up the equipment for the first test trial. The OPORDER described the scenario that the student was to follow, i.e., check-point times, attack time, etc. The OPORDER was similar to one that a TC might receive in the field, except that, in order to speed up the game, the student was asked to advance the clock to attack time when all battle positions had been prepared, rather than waiting for real time to elapse. The OPORDER was designed to incorporate into the game those factors identified by the task analysis described above that TCs frequently fail to perform satisfactorily. Basically, the OPORDER required the TC to prepare positions to defend a bridge, then proceed to prepare positions to attack an anti-tank guided missile. The scenario was developed to permit an evaluation of decision-making factors encountered in commanding a tank. The scenario was designed to provide sufficient daylight time for preparing battle position for nighttime operations. The game ended when the target was destroyed. The OPORDER is shown in Appendix E. The same OPORDER was used for the two different maps, as each student performed the following three trials.

Trial 1. In order to help the student understand the OPORDER, it was read aloud by the test administrator, while Map A was displayed. After the equipment was set up, the student was asked to point to the target to ensure that he knew where to attack. The student then proceeded to follow the OPORDER. There were several obvious bad routes and battle positions on Map A, but no obvious best route.

Trial 2. While the test administrator was setting up the equipment for the second trial, the student was asked to reread the OPORDER to enhance familiarity and understanding of the OPORDER. Map A was again displayed and the student proceeded to carry out the OPORDER.

Trial 3. Map B was displayed and the student executed the same OPORDER. There is an obvious best route on the map but also a "red herring," to attract a TC who is predisposed to go "right down the middle." The student was asked to read the OPORDER and was asked, prior to the start of the trial, to point to the target and to the particular bridge that was to be defended.

Results

Two kinds of analyses were performed: factor analysis of the Tank Commander game measures, and three-group multiple discriminant functions (the Tank Commander measures against the three criterion groups).

The factor analysis was carried out on 13 Tank Commander game measures listed in Table 4. The intercorrelations among the 13 measures were calculated using pair-wise deletion of missing values and then factored by principal components, followed by varimax rotation to simple structure. Five factors had eigenvalues greater than unity. Table 5 contains all loadings on the rotated factors equal to or greater than .50. Three of the factors are accounted for by LOS, LOSX, and HAZ for each of the three trials. The fourth factor is identified by the average high probability on trials 1 and 2, and the fifth factor pairs (inversely) AVE on the third trial with the total number of stakes placed for all three trials. The factor structure in Table 5 is encouraging in that the data are strongly patterned, logical, and loadings are high.

Students were instructed to advance the clock to 2300 when all battle positions had been prepared for the nighttime attack. In preparing the software it was assumed that a TC would not prepare positions while in line-of-sight of the target, however, a few students did. Whenever a student advanced the clock while in line-of-sight, misleading data were recorded for line-of-sight measures. For trial 1 there were six misleading scores. For trial 2 there were five misleading scores, and for trial 3 there were two misleading scores. In calculating the discriminant function one cannot utilize pair-wise deletion of missing values, but must delete all of the scores for a given student if there is a missing value on any one of them. It was necessary to delete these data from analysis, therefore, complete data were not available for all 40 officers. By using the third trial for the primary discriminant analysis, data for the maximum number of students were maintained. The result was that for the first discriminant analysis, trial 3, the LOS, LOSX, and HAZ measures were used, and it was only necessary to drop two students from the analysis.

Table 4
Means and Standard Deviations for the AOB Class for Various Tank Commander Game Measures

Measure/Trial		N	Mean	SD
LOS	1	34	164.8	113.1
LOS	2	34	180.2	258.6
LOS	3	38	102.3	53.0
LOSX	1	34	622.6	501.1
LOSX	2	34	620.6	1012.4
LOSX	3	38	317.8	128.2
HAZ	1	34	23.2	64.5
HAZ	2	33	54.2	225.0
HAZ	3	38	17.9	35.7
AVE	1	40	0.53	0.19
AVE	2	40	0.55	0.15
AVE	3	40	0.70	0.14
Total Stakes		40	5.90	5.14

Table 5
Factor Loadings for Selected Tank Commander Measures

Measure	I	II	III	IV	V
Trial 1					
LOS	.74				
LOSX	.99				
HAZ	.91				
Trial 2					
LOS		.87			
LOSX		.97			
HAZ		.95			
Trial 3					
LOS			.93		
LOSX			.95		
HAZ			.90		
Trial 1					
AVE				.87	
Trial 2					
AVE				.57	
Trial 3					
AVE					.90
Total number of stakes					-.80

Table 6 contains the means for the three groups on each of the included seven measures. Group 1 is "upper," Group 2 is "middle," and Group 3 is "lower." Significance levels (from a one-way analysis of variance for the three groups) are presented in the right-most column.

Table 6
Group Means and Significance Levels for Selected Tank Commander Measures

Measure	Group 1 N=14	Group 2 N=12	Group 3 N=12	Significance
Trial 1				
AVE	0.56	0.49	0.54	.682
Trial 2				
AVE	0.56	0.50	0.58	.488
Trial 3				
AVE	0.74	0.67	0.66	.273
LOS	119.4	106.5	78.2	.135
LOSX	349.2	366.0	233.1	.016
HAZ	27.6	23.5	1.0	.133
Total No. of Stakes	5.21	6.50	6.08	.825

The analysis was carried out using the seven measures in Table 6, all three criterion groups, and the Wilks' lambda criterion. The analysis allowed for the possibility that the three groups may not be adequately described in terms of a single dimension. Four measures were selected as more likely contributing to the predictive relation than not contributing to it. They were: trial 3 LOS, LOSX, HAZ, and AVE. The canonical correlation for the first discriminant function is .615 (significant at the .01 level). Wilks' lambda after removing the first function is .869 (significant at the level of .195). In short, there is some suggestion of a second function here, but it is not significant. The main result is a substantial predictive relationship (.615), albeit in the unanticipated direction, i.e., poor students performed well on the test, while good students did not do well on the test.

A discriminant analysis was conducted using the 13 measures in Table 5. Eleven students were lost to the analysis because of missing values. In this second analysis five measures were selected in addition to the four surviving from the first analysis. These were: trial 1 LOS; trial 2 LOS, LOSX, and HAZ; and the total number of stakes placed. The standardized coefficients for these new measures were small. Only one of them (for total number of stakes) was larger than the smallest coefficient for one of the four original measures (for trial 3 AVE). The first discriminant function in this second analysis has a canonical correlation of .847 (significant at the .005 level). Wilks' lambda after removing this function is .625 (significant at the .242 level), larger than in the first analysis. This result suggests that the second function is probably not there.

This higher canonical correlation (.847 versus .615) should be regarded with considerable suspicion. First, because the number of students was substantially reduced. Second, because of the six new measures considered for possible selection five of them were, in fact, selected. Third, the new variables have small standardized coefficients. Taken together, these three considerations suggest that this second canonical correlation capitalizes heavily on chance.

The OPORDER dictated that stakes be placed in positions to protect the bridge and at positions to prepare for a nighttime attack on the target. These positions were evaluated and scores were assigned according to the value (from 0 to 4) of each staked position. Table 7 presents the means and standard deviations of total stakes placed for the three groups of AOB students. Stake scores were not found to be related to class ranking by the instructors.

Table 7
Means and Standard Deviations for the Bridge and Target Staking Scores in the Upper, Middle, and Lower AOB Groups

		Bridge		Target	
Group	N	Mean	SD	Mean	SD
Upper	14	18.0	12.8	15.9	12.1
Middle	13	19.4	19.7	14.1	16.8
Lower	13	21.0	19.7	12.5	16.2

Discussion

The Tank Commander game was designed on the basis of a task analysis carried out with respect to enlisted TCs and using recent MILES exercises at the National Training Center as points of reference. The data in experiment 1 concern officers just completing basic armor training and using the Ten-Day War as a point of reference. There are two differences here: first, between enlisted men and officers and, secondly, between an exercise early in a tanker's career (the Ten-Day War) and one that normally takes place much later (MILES exercises at the National Training Center). It can be seen in Table 5 that highly ranked AOB students performed on the Tank Commander game in an "aggressive" manner, in that they were distinguished from the other two groups mainly by having a higher hit probability at the end (closer to the target when they fired), and placed fewer stakes. The students ranked lower in the class performed on the Tank Commander game in a manner that would be described as "longheaded" from the task analysis, however, training officers may see this behavior as "tentative" or overly cautious. It is likely that armor officers may be encouraged at the outset of their careers to behave in an agressive, take-charge manner and are rated that way. On the other hand, it may be that older, more experienced enlisted TCs are judged mainly on longheadedness (the results of the task analysis suggests that is a quality that some of them lack). Finally, it may be that what is considered optimal behavior in a tank officer may change in the course of his career. In the beginning aggressiveness may be at a premium, but later on it may be longheadedness.

Experiment 2

Method

Subjects

The performance of 30 enlisted men was evaluated. The enlisted men were student TCs in the Basic Non-Commissioned Officer Course (BNCOC). Two Military Occupational Specialties (MOSs) were represented: 19K (M1 tank) and 19E (M60 tank). Students in three classes were tested prior to commencement of training. Seven students were in training to become commanders of the M1 tank (MOS 19K). Twelve students were in training to become commanders of the M60A1 tank (MOS 19E), and 11 students were in training to become commanders of the M60A3 tank (MOS 19E). All seven of the M1 students were E5s, and of the M60 students, 16 were E5s and seven were E6s.

Materials

A series of tests, three microprocessor-based and five paper-and-pencil, were selected for administration as prospective selection tests. The tests were selected because they were expected or were known to have one or more attributes (to a large or small extent) that may be classed under eight validity criteria: predictive, metric, task, construct, content, concurrent, face, and factor. Rating and ranking forms were developed for the collection of criterion data.

The two video games were selected for quite different reasons. ACM is the best studied video game we have and is known to predict performance in a flight training simulator (Lintern and Kennedy, 1982); therefore, it might also predict TC performance. To the extent that TC performance can be expected to require more than the minimum of manual control skill, ACM performance should be predictive. It may be, however, that tracking and manual control skills would be more predictive of gunner and driver positions rather than tank commander. Touch-Me is a video version of digit span and was included as a measure of short-term memory, which also seems to be involved implicitly in tank commandering.

There were three testing stations (rooms) and three test administrators for each class of BNCOC students. One station was for paper-and-pencil tests (about 45 minutes). One station was for the ATARI games, Combat (ACM) and Touch-Me (about 45 minutes). The other station was for the Tank Commander game described in Experiment 1 (about 45 minutes). Total testing time, including instructions, for each student was approximately two hours and 15 minutes.

Predictors. The first three tests below, Hidden Patterns, Map Planning, and Paper Folding, were from the ETS Kit of Factor-Referenced Cognitive Tests and the descriptions are from the manual (Ekstrom et al., 1976). The ETS tests were originally designed to be administered in two parts to provide the capability to compute reliabilities. The five paper-and-pencil tests were:

1. Hidden Patterns. The factor involved in the Hidden Patterns test is the subject's "ability to hold a given visual percept or configuration in mind so as to disembed it from other well defined material" (Ekstrom et al., 1976, p. 19). In the Hidden Patterns test "each item consists of a given geometrical pattern in some of which a single given configuration is embedded. The task is to mark, for each pattern, whether or not the configuration occurs" (Ekstrom et al., 1976, p. 20).

2. Map Planning. The factor involved in the Map Planning test is the subject's "speed in exploring visually a wide or complicated spatial field" (Ekstrom et al., 1976, p. 155).

3. Paper Folding. The Paper Folding test measures "the ability to manipulate or transform the image of spatial patterns into other arrangements" (Ekstrom et al., 1976, p. 173).

4. Grammatical Reasoning. The Grammatical Reasoning test (Baddeley, 1968) purports to measure "higher mental processes." The measures of grammatical reasoning tend to correlate highly with a subject's verbal intelligence and are useful for differentiating people of disparate ability. The test consists of 32 complex statements concerning five grammatical transformations on statements about the relationships between the letters A and B. The subject responds "true" or "false" depending upon the verity of each statement.

5. Math Test. The Math Test consists of vertical addition problems involving three 2-digit integers. This test measures a subject's speed and accuracy for performing basic mathematical functions.

The three computer-based tests, each individually administered, involved interaction between the student and the computer, i.e., in each case the action of the student determined the response of the computer and vise versa. Two of the tests, Combat (ACM) and Touch-Me, are commercially available and the third, Tank Commander, was developed specifically for this investigation.

1. ACM. The ATARI Combat cartridge contains 25 games. The game selected for use in this study was game number 23 on the cartridge and involved jet-fighters in open skies with guided missiles, and was set at the novice skill level. A further description of the ACM task can be found in ATARI (1977). The game is referred to in the open literature as air combat maneuvering (ACM). It was played on a 24" color TV. Each student played 10 games (two minutes and 16 seconds each) against the computer. The score was the average number of hits over ten trials.

2. Touch-Me. Touch-Me is an ATARI product which tests short-term memory by requiring the subject to repeat, via a four-cell keyboard, stimuli presented visually and auditorially by the machine. The test apparatus is an inexpensive, rugged, highly portable (hand-held) battery-powered microcomputer. The objective of the game was to recall a series of tones and lights displayed at a rate of approximately .8 seconds each. The task was to key, in order, the appropriate buttons. Each time there was a correct response the device added another item and presented the sequence again. The game was set at the most difficult level and after

three misses a new trial was started. Each subject played the game for 15 minutes. Scores for all trials were recorded and the final score was the average of the three highest trials.

3. Tank Commander. This test is described in Experiment 1 and Appendix A.

Criterion Measures. The criterion variables (job dimensions) of interest for this investigation were those identified by the task analysis and the Basic Non-Commissioned Officer Course (BNCOC) cadre. The BNCOC cadre were experienced TCs. These job dimensions were not necessarily the most important aspects of commandering a tank, but rather those factors poor tank commanders frequently fail to perform satisfactorily during field exercises.

The BNCOC cadre observed the students daily in the classroom and in the field and were aware of the capabilities and limitations of each student. The cadre met the criteria identified by Murphy (1980, p. 48) as prerequisite for participation in the development of rating scales in that: they were supervisors who shared a common core of experience and of values concerning behavior on the jobs they rated; they had firsthand knowledge of what behaviors occur on the job, and of those behaviors which lead to acceptable performance; and they were familiar with organizationally peculiar language which reduces ambiguities in scaling. According to Fogli, Hulin, and Blood (1971), participation by supervisors, as developers and evaluators, should have a favorable impact on both validity and reliability of behaviorally anchored rating scales.

Two groups of SMEs were used to develop the rating scales. Procedures used for identifying job dimensions for the rating scales from the first group are described in the task analysis section. The first group was made up of the platoon leaders and platoon sergeants recently returned from field exercises. The second group used for development of the scales was the BNCOC instructors who were responsible for the training and assessment of the performance of the students. The objective of the behaviorally anchored rating scale developmental process was to break the tank commander job down into component parts, or various job dimensions, in order to describe the range of specific behaviors which the instructor cadre might reasonably expect to see a student exhibit. Then behavior within each job dimension was scaled from what was deemed by the cadre as behavior most representative of a good tank commander to those deemed to represent ineffective behavior. This procedure, whereby the consensus of a qualified group of raters is evaluated by a second equally qualified group, contributes to the reliability of the scale (Smith & Kendall, 1963).

Supervisory rating scales and ranking techniques were developed for the M1, A1, and A3 BNCOC students, and peer ratings for the A3 class. These are shown in Appendix J. The skill areas identified for assessment were derived from the results of the task analysis shown in Table 3 and modified by the BNCOC cadre. The five skill areas identified by the cadre were: threat avoidance, planning battle positions, movement after firing, conformance to doctrine, and resource management. Each student was rated, by supervisory cadre, upon performance in each of the five skill areas. The cadre were presented with four behaviorally anchored choices for each skill area.

The BNCOC cadre was asked to rank all students from most proficient to least proficient on their overall performance in the courses. Also, they were asked to estimate the percentile into which each student would fall, based upon their general experience with other TC classes. The peer rating scale administered to the A3 class only, took the form of assigning each student to the loader position and asking him to name the three students in his class that he would prefer to have as his TC on a hazardous mission.

The other criterion measure used was rankit. Rankit is determined as follows. If one draws a sample of _N_ subjects from a unit-normal population, the i _th_ subject in order of magnitude has an expected value expressed in standard deviations units. The rankit transformation carries each rank into its expected normal deviate as just described. The effect is "to spread out" the extreme ranks relative to those more toward the middle of the range, so that the step between the highest and the next highest subjects is substantially larger than the step between neighboring subjects around the median.

Procedures

The BNCOC students were administered five paper-and-pencil tests and three interactive computer tests. Each class was tested on the two days prior to reporting for BNCOC. Students were provided with a copy of the Privacy Act. Testing was conducted simultaneously at all three testing stations. At each station, the test administrator followed detailed written procedures, read the instructions to each student for each test, and timed the test. The following procedures were executed.

Paper-and-pencil instructions and procedures, as shown in Appendix F, were followed. Prior to administration of the paper-and-pencil tests, each student was asked to fill out a short questionnaire regarding experience playing home video games (see

Appendix G). At the ATARI station, each student was tested on the Combat and Touch-Me games. The instructions for Combat, shown in Appendix H, were read and the student then performed ten trials. Upon completion of ACM, each student was read the instructions for Touch-Me and then performed for 15 minutes. The instructions for Touch-Me can be seen in Appendix I. For the tank commander station, the procedures described in Experiment 1 were followed.

Results

The Tank Commander test advancing the clock in line-of-sight problem reported in Experiment 1 also existed in the present experiment. Therefore, it was impossible to obtain valid data on all of the measurements for each of the students. In the present experiment this problem was worsened by the small number of students in each group. Simply excluding a student who advanced the clock while in line-of-sight would have decreased the number of students to an unacceptable size. Therefore, instead of treating three separate scores (trials) for each Tank Commander task measure for each student, one score was calculated. In each case the score on a given measure is the average of however many valid trials were available, without regard to which trials they were.

The results from the M1 and M60A1 (called A1) classes were similar to each other in that in both cases the hypotheses were supported that poor students would do less well on selected measures from the Tank Commander test than would good students. On the other hand, the opposite was true for the M60A3 (A3) class tested later, therefore, the data from the first two classes were combined for analysis, while the M60A3 class data were treated separately. Means and standard deviations for the first two classes can be seen in Table 8.

Table 8
Means and Standard Deviations of Tank Commander Measures for the M1 and M60A1 Classes Combined

Measure	N	Mean	SD
LOS	19	278.0	157.8
LOSX	19	1041.5	821.2
LOSP	19	69.4	59.8
HAZ	19	121.8	117.5
AVE	19	0.6	0.2
Total Stakes	17	12.4	5.3

There are three criterion measures for each student: the overall rank of the student as evaluated by the training cadre and then rankit transformed, the sum of the five behaviorally anchored rating scales (BARS) answered by the training cadre on each student, and the percent rating of the student relative to the cadres' "general experience with other TC students". Correlations between all of the predictor and criterion measures are shown in Appendix K.

Tables 9, 10, and 11 present the correlations between the Tank Commander game and the criterion measures. The first and most obvious point is that the correlations involving LOS, LOSX, LOSP, and HAZ are all negative; the larger the Tank Commander measure the lower the criterion score. Second, these correlations are sizable. It should be noted, moreover, that unlike the correlation of .615 in the Experiment 1 AOB sample, for the relationship between the Tank Commander game measures and overall rank, the present correlations involve no capitalization on chance. These are not multiple but first-order correlations. Multiple correlations for such small groups would be inappropriate; the resulting values would be inflated. Third, the overall rankit score is as good as any of the criterion measures. Not much, if anything was gained by the BARS items, nor the peer ratings. Finally, it should be noted that the correlations for the two groups combined involving AVE and total number of stakes placed are not negligible. Their direction, moreover, is puzzling: the lower the probability of hitting the target when the student fires, and when the student places fewer stakes, the higher the criterion score.

Table 9
Correlations of Selected Game Measures With Overall Rank (Rankit Transformed), Summed BARS, and Percent Standing in the M1 Class (N=7)

Measure	Overall Rankit	Summed BARS	Percent Standing
LOS	-.62	-.77*	-.79*
LOSX	-.41	-.59	-.55
LOSP	-.48	-.65	-.61
HAZ	-.64	-.73*	-.79*
AVE	-.07	.10	.05
Stakes	-.29	-.36	-.40

* $p < .05$.

Table 10
Correlations of Selected Tank Commander Measures With Overall Rank (Rankit Transformed), Summed BARS, and Percent Standing in the M60A1 Class (N=12)

Measure	Overall Rankit	Summed BARS	Percent Standing
LOS	-.65*	-.43	-.50
LOSX	-.61*	-.29	-.39
LOSP	-.65*	-.33	-.42
HAZ	-.68**	-.37	-.43
AVE	-.57*	-.56*	-.59*
Stakes	-.52	-.78**	-.51

* $p < .05$. ** $p < .01$.

Table 11
Correlations of Selected Tank Commander Measures, Summed BARS, and Percent Standing in the M1 and M60A1 Classes Combined (N=19)

Measure	Overall Rankit	Summed BARS	Percent Standing
LOS	-.60**	-.42	-.47*
LOSX	-.53*	-.25	-.23
LOSP	-.57**	-.30	-.27
HAZ	-.63**	-.39	-.36
AVE	-.46*	-.28	-.31
Stakes	-.43	-.42	-.34

* $p < .05$. ** $p < .01$.

Table 12 contains correlations between overall rankit and each of the seven reference tests in the M1, M60A1, and combined classes. The two commercially available video games, Combat and Touch-Me, seem not to be appreciably related to overall rankit. Of the five paper-and-pencil tests, Hidden Patterns and Paper Folding are the best. They relate to overall rankit at about the same level as the line-of-sight measures do in the Tank Commander game.

Table 12
Correlations Between Overall Rankit (Rankit Transformed) and the Reference Tests in the M1 and the M60A1 Classes, Separately and Combined

Reference Test	M1 (N=7)	A1 (N=12)	M1 + A1
Mathematics	.29	.32	.28
Grammatical Reasoning	-.59	-.27	-.38
Hidden Patterns	.53	.63*	.58**
Map Planning	.43	.13	.23
Paper Folding	.55	.73	.66**
Combat	.03	.42	.27
Touch-Me	.29	.32	.31

* $p < .05$. ** $p < .01$.

Table 13 presents the correlations between Hidden Patterns and Paper Folding, and the line-of-sight measures of the Tank Commander game. The correlations are high, suggesting that both Tank Commander and the criterion involve substantial elements of spatial or perceptual judgment. On the other hand, the correlations are not so high as to preclude potential for unique contributions to overall rankit by both Tank Commander and the paper-and-pencil tests.

Table 13
Correlations Between the Tank Commander Game Measures and the Hidden Patterns and Paper Folding Tests, for the Combined M1 and M60A1 Classes (N=19)

Game Measure	Hidden Patterns	Paper Folding
LOS	-.73**	-.75**
LOSX	-.56**	-.61**
LOSP	-.57**	-.62**
HAZ	-.66**	-.70**
AVE	-.14	-.26
Stakes	-.42	-.39

** $p < .01$.

Table 14 presents the correlations calculated within groups among Hidden Patterns, Map Planning, Paper Folding, and student rank, i.e., rankit transformed, for all three groups (N=30). Paper Folding is by far the most predictive of the three tests (.618), with Hidden Patterns and Map Planning correlating much less strongly (.434 and .342, respectively). The multiple correlation between the three tests and overall student rank is .642. Hidden Patterns adds nothing to Paper Folding, despite the correlation of .434 between Hidden Patterns and the criterion: the reason being the high (.691) correlation between Paper Folding and Hidden Patterns. Map Planning does a little better, increasing the multiple correlation from .618, the value for Paper Folding alone, to .641. A linear composite consisting of Paper Folding and Map Planning does essentially as well as all three tests.

Table 14
Within Group Correlations Among Hidden Patterns, Map Planning, Paper Folding, and Overall Rank (Rankit Transformed) Over All Three Groups (N=30)

Measure	HP	MP	PF	SR
Hidden Patterns (HP)	--	.12	.69**	.43*
Map Planning (MP)		--	.30	.35*
Paper Folding (PF)			--	.62**
Student Rankit (SR)				

* $p < .05$. ** $p < .01$.

Stakes represent planned battle positions and the positions were assigned scores, or values, according to whether they were placed to protect the bridge, or as planned positions for attacking the target. The results for the staking scores are shown in Table 15. The bridge staking score shows no association with rankit transformed and the target staking score shows a non-significantly negative correlation with student rankit. The conclusion is that the staking scores do not relate to TC performance as evaluated by the training cadre.

Table 15
Correlations Between Student Rank (Rankit Transformed) and the Bridge and Target Staking Scores in the M1, A1, and A3 Classes

Class	N	Bridge	Target
M1	7	.04	-.31
A1	12	.05	-.73**
A3	11	.02	.18

** $p < .01$.

For the BNCOC M60A3 class the correlation between overall rankit score and LOS was +.32, that is, the better the overall rankit the longer a student spent in LOS, in this respect like the

AOB students in Experiment 1 and contrary to our fundamental hypothesis. The correlations between the other Tank Commander measures and rankit were in the same direction as LOS: precisely opposite from the results obtained on the other two BNCOC classes. The mean rankit score for the E6s was +.24, while for the E5s the mean was -.11. The estimated within group standard deviation was .93, yielding $t=.83$, ($p = .58$, with 21 degrees of freedom). Thus, the E6s did better on rankit, but not significantly.

The reference tests related in the same ways for the M60A3 class as they did in the earlier samples. Table 16 contains the correlations between the seven reference tests (five paper-and-pencil and two video) and the rankit score in the three samples.

Table 16
Correlations Between Overall Rank (Rankit Transformed) and the Reference Tests in the M1 and M60A Classes and Overall Within Classes

Reference Test	M1 (N=7)	A1 (N=12)	A3 (N=11)	Overall
Mathematics	.29	.32	.18	.24
Grammatical Reasoning	-.59	-.27	-.16	-.31
Hidden Patterns	.53	.63*	.08	.43*
Map Planning	.43	.13	.64*	.35*
Paper Folding	.55	.73**	.53	.62**
ACM (video)	.03	.42	.69*	.46**
Touch-Me (video)	.29	.32	.61*	.43*

* $p < .05$. ** $p < .01$.

In Table 16 the relationships are consistent throughout and, with the exception of Mathematics and Grammatical Reasoning, sizable. These are first-order correlations, therefore, there is little possibility of capitalization on chance.

Discussion

The major finding resulting from this experiment is that three short, standardized tests of spatial ability, Paper Folding, Map Planning, and Hidden Patterns, correlated ($R = .642$) with TC performance as rated by the BNCOC cadre. Two of the tests, Paper Folding and Map Planning, correlated almost as well, .641. In a cross validation sample, one would expect these correlations to drop. How much is difficult to say. If one corrects the obtained values for shrinkage, the value of .642 for all three tests shrinks to .587, and the value of .641 for Paper Folding and Map Planning alone shrinks to .606. Shrinkage, however, corrects for capitalization of sample variations only, not test selection, and there is some of the latter also involved in the obtained values.

The three spatial tests were selected from a total of seven tests, not counting the Tank Commander test. Paper Folding and Map Planning were selected after the fact as the best combination of two tests. It is reasonable, therefore, to expect some shrinkage above and beyond the values corrected solely for capitalization on sampling variations. A likely figure might be .55.

If cross validation did, in fact, yield a value as high as .55, the result would be outstanding. In his well known survey of predictive validities for job performance, Ghiselli (1966) found an average value of .19. For training criteria the average is higher, but not greatly so. Predictive validity ranges in the neighborhood of .20. Even, therefore, in a validation sample, .55 is promising indeed.

The interpretation of this result is uncertain. In the first place, these tests are not job samples. In a general way one can say that spatial abilities are involved in tank commandering, but there is no obvious common content between the three spatial tests and what a TC does. If spatial ability is the link making for predictive validity, it is manifested differently in the tests than in the criterion. In fact, however, the link may be broader than spatial ability. Spatial tests are the best known indicators of right-brain function, which includes other things than spatial ability, for example, sense of direction, the ability to visualize, and certain of Piaget's conservation tasks possibly creativity, and a quick and ready intuitive grasp of concrete situations (McGee, 1977). Spatial ability seems also to be linked with hormonal influences, especially at puberty (Hier & Crowley, 1982; Nyborg & Nielsen, 1981). It seems possible, therefore, that the link between the three spatial tests and TC performance is right-brain function generally, rather than spatial ability in a narrow sense. The negative correlations between Grammatical Reasoning and TC performance would reinforce this interpretation. The reasons for including Grammatical Reasoning were not completely for research purposes or just for promoting understanding. The metric properties were very important but also it was felt that the factor purity of the tests would aid in interpretation of the nature of the TC's task. Moreover, it was predicted at the outset that the likely relationship to performance on the Grammatical Reasoning test would be negatively related to performance as a TC. Specifically, it may be that the work that TCs might be expected to perform is associated with so-called right-brain activities, whereas grammatical reasoning, being a largely verbal analogical reasoning, is the type of function that one would ordinarily associate with left-brain activity (Andreassi & Juszczak, 1983; Tipton & Mohler, 1983). If this hypothesis concerning right-brain function has merit, a logical way to proceed would be to augment the spatial tests with measures of field dependence/independence, performance, and tachistoscopic perceptual tests.

General Discussion

The major finding in the work being reported, aside from the strong correlations for two paper-and-pencil tests, concerns the Tank Commander video task. The results obtained do not seem to indicate that the task is simply irrelevent to TC performance. The problem seems rather to be a confusion between longheadedness and lack of aggressiveness or, conversely, between shortheadedness and aggressivity. No one should be surprised by this confusion since it exists in many contexts other than the TC task. It is often difficult to know whether a given action reflects timidity or foresight, a healthy aggressiveness of rashness. The problem with the TC task is how to disentangle behaviorally the two counter veiling trait/interpretations.

One possibility is to redesign the Tanker Map so as to allow behaviors that would not support any interpretation but stupidity. Even if the behaviors were infrequent, as long as they occurred at all, they might allow us to identify some definitely non-thinking TCs. The first map, for example, that Mr. Schafer designed contained a "corral" from which it was impossible to exit except through a narrow passage that opened up into an area directly exposed to enemy fire. This map was not used, in part because it seemed that no one would be dumb enough to enter the corral. This may have been a mistake; perhaps some subjects would have entered it and, if they had, the fact might have been predictive of poor performance as a TC.

A second possible approach to redesigning the Tank Commander task is to examine a subject's performance much more closely than was done. The analysis would be based on a kind of script (Abelson, 1981). In some instances, a subject may accumulate time in LOS in a bold dash across an enemy line of fire to reach a strong attacking position. Such a dash might be scored as aggressiveness, not failure to foresee adverse consequences. In other instances a subject may accumulate time in LOS by traveling within range of enemy guns when other routes are available. In such a case the exposure to enemy fire would be thoughtless.

This second approach would depend not so much on redesigning the Tanker Map, as on a more sophisticated means of scoring subject performance. Since this scoring would presumably be automatic, the approach implies a scoring algorithm capable of distinguishing between a "bold dash" and "thoughtless traveling." Such an algorithm implies a high order of sophistication and ought not to be underestimated.

The best continuation of the present work would follow both of these approaches to redesigning the Tank Commander game and would also, one hopes, pursue the strongly positive findings obtained with the spatial tests. Still another possible continuation would be to develop the Tank Commander game as a training device. Many observers, especially tankers, have commented on the task's similarity to sand tables and terrain boards and pointed out the advantages that a developed dynamic video-computer task would have over them.

References

Abelson, R.P. (1981). Psychological status of the script concept. American Psychologist, 36, 715-729.

American Psychological Association (1974). Standards for educational and psychological tests. Washington, D.C.: Author.

Andreassi, J.L., & Juszczak, N.M. (1983). Brain responses and information processing IV: Line discrimination, color, perception, masking, and shape discrimination. Washington, D.C.: Air Force Office of Scientific Research (NL).

ATARI, Inc. (1977). Combat game program instructions. Sunnyvale, CA: ATARI, Inc., Consumer Division, (C011402-01).

Baier, D.E. (1947). Selection and evaluation of West Point Cadets. American Psychologist, 2, 325.

Baker, J.D. (1981). Principles for design for educational games and simulations. Paper presented to the 1981 Summer Simulation Conference, Washington, D.C.

Baddeley, A.D. (1968). A three minute reasoning test based upon grammatical transformation. Psychonomic Science, 10, 341-342.

Bersoff, D.N. (1981). Testing and the law. American Psychologist, 36, 1047-1056.

Bittner, A.C., Carter, R.C., Harbeson, M.M., & Krause, M. (1983). Repeated measurement of human performance: The good, the bad, and the ugly. New Orleans, LA: Naval Biodynamics Laboratory, in preparation.

Bittner, A.C., Carter, R.C., Krause, M., Kennedy, R.S., & Harbeson, M.M. (1982). Performance tests for repeated measures: Moran and computer batteries. Proceedings of the 26th Annual Meeting of the Human Factors Society, Seattle.

Bolger, J.P. (1981). Cognitive retraining: A holistic approach. Unpublished manuscript.

Buros, O.K. (1972). The seventh mental measurements yearbook. Highland Park, NJ: The Gryphon Press.

Campbell, J.P., Dunnette, M.D., Arvey, R.D., & Hellervik, L.V. (1973). The development and evaluation of behaviorally based rating scales. Journal of Applied Psychology, 57, 15-22.

Carter, R.C., Kennedy, R.S., & Bittner, A.C. (1981). Grammatical reasoning: A stable performance yardstick. Human Factors, 23 (5), 587-591.

Carter, R.C., & Shannon, R.S. (In press). Repeated measures of search tasks. New Orleans, LA: Naval Biodynamics Laboratory Technical report.

Doll, R.E. (1963). Officer peer ratings as predictor of failure to complet flight training. Aerospace Medicine, 34, 130-131.

Drucker, E.H. & O'Brien, R.E. (1981). Mission - based analysis of armor training requirements (Vol. 1: Final Report) (HumRRO FR-MTRD (KY)-81-2). Fort Knox, KY: Human Resources Research Organization.

Eaton, N.K. (1978). Performance motivation in armor training (Technical Paper 291). Fort Knox, KY: ARI, AD 064274.

Eaton, N.K., Bessemer, D.W., & Kristiansen, D.M. (1979). Tank crew position assignment (Technical Report 391). Fort Knox, KY: ARI.

Eaton, N.K., Johnson, J., & Black, B.A. (1980). Job samples as tank gunnery performance predictors (Technical Report 473). Alexandria, VA: U.S. Army Research Institute.

Eaton, N.K., & Whalen, D.E. (1980). Accuracy of tank gunnery scoring (Working Paper FKFU 80-4). Fort Knox, KY: ARI.

Ekstrom, R.B., French, J.W., Harmon, H.H., & Derman, D. (1976). Manual for kit of factor-referenced cognitive tests (Office of Naval Research Contract N00014-71-0117, Project Designation NR 150 329). Princeton, NJ: Educational Testing Service.

Enneis, W.H. (1970). The black man in the world of work. Professional Psychology, 1, 436.

Flanagan, J.C. (1954). The critical incident technique. Psychological Bulletin, 51, (4).

Fleishman, E.A. (1975b). Taxonomic issues in human performance research. In W.T. Singleton and P. Spurgeon (Eds.), Measurement of human resources. New York: Halstead Press.

Fleishman, E.A. (1975a). Toward a taxonomy of human performance. American Psychologist, 30, 1127-1149.

Fogli, L., Hulin, C.L., & Blood, M.R. (1971). Development of first-level behavioral job criteria. Psychological Bulletin, 55, 3-8.

Ghiselli, E.F. (1966). The validity of occupational aptitude tests. New York: Wiley.

Glaser, R., & Bond, L. (Eds.), (1981). Testing: Concepts, policy, practice, & research (special issue). American Psychologist, 36.

Goldin, S.E., & Thorndyke, P.W. (1981a). An analysis of cognitive mapping skill (ARI Contract MDA903-79-C-0549). Santa Monica, CA: Rand Corporation.

Goldin, S.E., & Thorndyke, P.W. (1981b). Spatial learning and reasoning skill (ARI Contract MDA903-79-C-0549). Santa Monica, CA: Rand Corporation.

Greenstein, R.B., & Huges, R.G. (1977). The development of discriminators for predicting success in armor crew positions (Research Memorandum 77-27). Alexandria, VA: U.S. Army Research Institute.

Guion, R.M. (1979). Principles of work sample testing: I A non-impirical taxonomy of test uses (TR-79-A8). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD AO-72446)

Guion, R.M., & Ironson, G.H. (1979). Principles of work sample testing: IV Generalizability (TR-79-A11). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD AO-71083)

Harari, O., & Zedeck, S. (1973). Development of behaviorally anchored scales for the evaluation of faculty teaching. Journal of Applied Psychology, 58, 261-265.

Hier, D.B., & Crowley, W.F., Jr. (1982). Spatial ability in androgen-deficient men. New England Journal of Medicine, 306, 1202-1205.

Hughes, R.G. (1976a). Development of predictors for task gunner. Unpublished manuscript.

Hughes, R.G. (1976b). Development of predictors for tank gunner. Unpublished manuscript (revised).

Hunter, D.R., Maurelli, V.A., & Thompson, N.A. (1977). Validation of a psychomotor/perceptual test battery (AFHRL-TR-77-28). Lackland Air Force Base, TX: Air Force Human Resources Laboratory, Personnel Research Division. (NTIS AD-AO-445-25)

Imhoff, D.L., & Levine, J.M. (1981). Perceptual-motor and cognitive performance task battery for pilot selection (AFHR-TR-80-27). Advanced Research Resources Organization, Washington, D.C., & Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Jones, M.B. (1972). Economic consequences of low IQ. Social Welfare Forum.

Jones, M.B. (1973). Moderated regression and equal opportunity. Educational and Psychological Measurement, 33, 591-602.

Jones, M.B. (1974). Regressing group on individual effectiveness. Organizational Behavior and Human Performance, 11(3).

Jones, M.B. (1981). Further studies on stabilization with practice in performance tasks (Final Report, Contract No. N0023-80-M-5389). New Orleans, LA: U.S. Naval Biodynamics Laboratory.

Jones, M.B. Kennedy, R.S., & Bittner, A.C., Jr. (1981). A video game for performance testing. American Journal of Psychology, 94, 143-152.

Jones, T.N., & Doll, R.E. (1974). Peer ranking as criterion measures for initial acquisition of targets (Aerospace Psychology Technical Memorandum 74-2). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Kennedy, R.S., Andrews, D.A., & Carter, R.C. (1981). Performance evaluation tests for environmental research (PETER): A microcomputer game as a memory test. Preprints of the 52nd Annual Scientific Meeting of the Aerospace Medical Association, San Antonio.

Kennedy, R.S., Bittner, A.C., Jr., Harbeson, M., and Jones, M.B. (in press). Television computer games: A 'new look' in performance testing. Aviation, Space, and Environmental Medicine.

Kennedy, R.S., Bittner, A.C., Jr., and Jones, M.B. (1981). Video-game and conventional tracking. Perceptual and Motor Skills, 53, 310.

Klemp, G.O., Jr., Munger, M.T., & Spencer, L.M., Jr. (1977). Analysis of leadership and management competencies of commissioned and noncommissioned officers in the Pacific and Atlantic fleets (Final Report, Task Planning Document E6-33 on Contract N00600-75-D-0038). Boston, MA: McBer and Company.

LaGiapa, J.J. (1947). Analysis of relative weights assigned to OCS predictors and OCS grades (Technical Bulletin 61-3). Washington, D.C.: Bureau of Naval Personnel.

Lintern, G., & Kennedy, R.S. (1982). A video game as a covariate for the carrier landing research. Proceedings of the 8th Psychology in the DoD Symposium (USAFA-TR-82-10). Colorado Springs, CO: USAF Academy, 311-315.

Lynch, W.J. (1981a). TV games as therapeutic intervention. Paper presented to the American Psychological Association, Los Angeles, CA.

Lynch, W.J. (1981b). A guide to Atari videocomputer programs for rehabilitation settings (revised). Unpublished paper.

McCormick, E.J. (1979). Job analysis: Methods and applications. New York: AMACOM.

McGee, M.G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological inferences. Psychological Bulletin, 86, 889-918.

Melton, A.W. (Ed.) (1947). Apparatus tests. Washington, D.C.: U.S. Government Printing Office.

Messick, S. & Jungblut, A. (1981). Time and method in coaching for the SAT. Psychological Bulletin, 89, 191-216.

Miller, R.B. (1962). Task description and analysis. In R.M. Gagne (Ed.). New York: Holt, Rinehart, Winston.

Murphy, J.W. (1980). Use of behaviorally anchored rating scales (BARS) to complement the management by objectives (MBO) and fitness report components of the Marine Corps performance evaluation system. Master's thesis, Military Art and Science, U.S. Army Command and General Staff College, Fort Levenworth, KS.

Nyborg, H., & Neilsen, J. (1981). Spatial ability of men with karyotype 47, XXY, 47, XYY, or normal controls. In W. Schmidt & J. Neilsen, Human behavior and genetics. Amsterdam: Elsevier/North Holland Biomedical Press.

Rigby, M.K., & Ossorio, E.D. (1959). Sociometric status and attitude characteristics of successful and unsuccessful Wave recruits (Technical Report No. 11). St. Louis, MO: St. Louis University.

Schmidt, F.L., Greenthal, A.L., Hunter, J.E., Berner, J.G., & Seaton, F.W. (1977). Job sample vs. paper-and-pencil trades and technical tests: Adverse impact and examinee attitudes. Personnel Psychology, 30.

Schwartz, D.J. (1977). A job sampling approach to merit system examining. Personnel Psychology, 30.

Simutis, Z.M., Baker, J.D., Bersh, P.J., & Alderman, I.N. (1979). An empirical validation of the effectiveness of a computerized game to teach troubleshooting. Paper presented to the American Educational Research Association, San Francisco.

Smith, P.C., & Kendall, L.M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.

Tipton, D.A., & Mohler, S.R. (1983). The Wright brothers: A personality profile. Aviation, Space, and Environmental Medicine, 54(6), 560-562.

Thomas, S.A., & Sternberg, J. (1964). Construction of an experimental selection battery for Armor systems. Army Research Institute Research Memorandum 64-2.

Thomas, J.O., & Weltman, G. (in press). A low-cost portable part-task training device for skilled performance training. Proceedings of the Second Interservice/Industry Training Equipment Conference. Salt Lake City, Utah.

Thorndyke, P.W., & Goldin, S.E. (1981). Ability differences and cognitive mapping skill (ARI Contract MDA903-79-C-0549). Santa Monica, CA: Rand Corporation.

Underwood, B.J., Borach, R.F., & Malmi, R.A. (1977). The composition of episodic memory (ONR Contract No. N00014-76-C-0270). Evanston, IL: Northwestern University. (AD-A0-40696)

Wagner, L.C., Jr. (MG). (1981). Commander's hatch: The basic NCO course. Armor.

Williams, S.B., & Leavitt, J.J. (1947). Group opinion as a predictor of military leadership. Journal of Consulting Psychology, 11, 283.

Appendix A

Description of the Tank Commander Video Game

Tank Commander Video Game

This video game, or test, was developed specifically for this experiment. The test is a simulation of tactical battlefield situations designed to evaluate the decision-making performance of a commander of a tank. It has the following features.

Hardware

The following equipment is required: an APPLE II plus 48k computer; two disk drives; a MIMCO STICK which includes a joystick for driving the tank in the moving mode or advancing the clock in the parked mode, a toggle switch for selecting parked or moving mode and three buttons for firing, staking, or advancing the clock; a SUPERCLOCK in slot 4 (or a CCS clock in slot 5); and a color monitor. The program is written in Pascal and a language interface card is required.

Software

The game consists of a colored map type display, on the monitor, of various terrain features such as mountains, hills, roads, rivers, etc. The software package consists of three programs, TANKER, TANKERMAP and STOREMAP. TANKERMAP and STOREMAP are used to generate and store the map, start times, locations and other information for the TANKER program. TANKERMAP and STOREMAP are operated by the person setting up the simulation. A map consists of 60 horizontal squares by 40 vertical squares. The squares are set up to correspond to 10 kilometers. Each square, or block, can be programmed to be blank, colored (representing terrain features), in line-of-sight to the target (T LOS), in line-of-sight to the bridge (B LOS), have a probability (PROB) of 0 to 9, be an EDGE square (there is LOS between own tank and the target, but not the reverse, as if own tank is hidden in trees), or be a hazard (HAZ) square (within range of target weapon). Hazzards, edges, and probabilities are not visible during operation of the TANKER simulation.

TANKER is the main simulation program that is intended for use by the test participant. TANKER has two operational modes, Moving and Parked. In the Parked Mode the gun can be fired, positions staked, clock advanced, range to map objectives determined, or the tank can be put into the moving mode. In the Moving Mode, only the speed and direction of the tank can be changed. The Tank is displayed as the cursor at the center of a white cross (X) and is programmed to approximate a maximum speed of 90 kilometers per hour. In order to speed up the game, computer time is three times as fast as real time.

Information displayed on the left side of the screen indicates, from present tank position, whether or not there is line-of-sight (LOS) to the target, distance to target, presence or absence of LOS to the bridge, distance to bridge, whether or not present position has been staked, and in the parked mode whether or not the READY light is on, meaning that the present position of the tank is in LOS and within range of own weapons (2,000 meters). At the lower left corner of the screen there is an outline of a circle with a pointer inside the cirlce which indicates the direction that the tank is moving, while in the moving mode. At the bottom of the screen, beneath the map, time of day is displayed and, beneath time, a COMMAND statement indicating whether the tank is in the moving or parked mode and, if in the parked mode which options are available (fire, stake or advanced clock).

Performance measures. Student performance is assessed by the computer program taking the following measures and recording them in a data file on a disk.

LOS. The total number of seconds spent in line-of-sight to the target. Five seconds after a shot was fired from an EDGE block, that position was changed to LOS and HAZ.

LOSX. This is a cumulative score derived by calculating the number of seconds spent in each LOS block multiplied by 10,000 and divided by the distance (range in meters) between each block and the target.

LOSP. The cumulative number of seconds spent at any block in T LOS multiplied by the probability associated with each block.

HAZ. The number of seconds spent in hazard blocks.

STAKES. The number of stakes placed; the X Y coordinates of each stake; the time each stake was placed; the hit probability associated with each stake; and whether or not the stake was placed on a HAZ, EDGE or T LOS square.

PROB. The hit probability associated with the positions of each of the first three shots. The computer was programmed so that the target could never be destroyed by the first three shots, in order to determine how well the student planned and executed tactics and used terrain. The computer was programmed so that the target was randomly destroyed after the third, but no later than the seventh shot. Destruction of the target was a function of the computer program and not dependent upon the skill of the player, except that the tank must be in line-of-sight to the target and be within own weapon range.

AVE. The average probability of hit of the first three shots.

SHOTS. The X Y coordinates of the tank, and the time, when each of the first three shots were fired.

Final Location Staked. An indication of whether or not the tank location was staked when the simulation ended.

Moved After Shot. An indication of whether the tank moved after the second shot.

Seconds Missed Assigned Time. The number of seconds between the assigned target attack time and the time when the first shot was fired.

Appendix B

Background and Video Experience: AOB Questionnaire

NAME: ______________________________

DATE AND TIME: ______________________

General Information

What is the source of your commission?

USMA ____ OCS ____ ROTC ____

DIRECT ____ OTHER ____

What is your army status?

RA ____ USAR ____ NG ____ OTHER (Specify) ____

How much tank experience do you have (in months)? _____

How frequently do you play video games?

Almost Daily ____ Weekly? ____

A few times a month ____ *Very Seldom or never* _____

Which video games? ____________________________

PT: 5540a

Appendix C

Instructions and Procedures for Tank Commander Video Game

Instruction for Administration of the Tank Commander Game

1. Each subject will play the game three times, after he has completed the demonstration trial. You will need the following disks:

   a. System Disk (Tank1, Drive #1). Student performance data filed on this disk.

   b. Demonstration Map (Tank2, Drive #2)

   c. Map A (map labeled Blasche, Tank2, Drive #2)

   d. Map B (map labeled Laura, Tank2, Drive #2)

2. To get the game started, follow the "Wise Owl Workshop" instructions for TANK2:TANKER. Insert the Tank1 Disk into Drive #1 and the Tank2 Demonstration Disk into Drive #2.

   a. Read the Instructions to the subject. Inform the subject that he can feel free, at any time to ask questions about the operation of the game, i.e., "How do I place a stake, advance clock, etc."

   b. When you start the Demo game, do not ENTER "ID=" information: just hit the return key. This will cause data to be stored as .TEXT. Occasionally, as you need space for data storage, delete the Demonstration data by following Wise Owl Workshop FILER instructions as follows:

You Enter	Screen displays
R	Remove?
.TEXT	Update?
Y	.TEXT Removed

3. After the demo game has been completed, give the student his OPORDER to read while you remove the Demo disk and insert the Blasche Map (TANK2) into Drive #2. When you have the game ready to run (Tank1 in Drive 1, Tank2 in Drive 2, joystick calibrated by adjusting controls above and on the right side when necessary), you must perform the following operations:

   a. ID 1= Enter nine (9) or fewer characters (letters and numbers only, no commas, etc.). In this case, Trail Number, Last Name and Initial (space permitting). The second ID- is the name of the map being used (Map A, "Blasche", or Map B, "Laura"), the third is Social Security Number, the fourth ID= is the class number and the fifth is the date. The screen will then look like this:

1 = 1 Doe J

2 = Blasche

3 = 123456789

4 = BNCOC XX

5 = 22 May 83

b. Next follow the instructions on the screen (hit any key or flip toggle switch, as directed), then:

(1) Read the OPORDER to the student.

(2) Ask the student if he has any questions on the operation of the game, i.e., staking, firing, advancing clock, etc. If he has questions on the OPORDER, tell him to reread it and take all of the time he needs.

c. After the target is destroyed at the end of the first trial, ask the student to read the OPORDER again while you are reinitializing the system for the second trial. The second trial is an exact replication of the first trial except it is only necessary to enter information on the first and second ID= lines, then hit Return. For trial two the first ID= entry wll be 2 followed by name, and the second line ID= entry is Blasche.

d. At the end of the second trial, ask the student to reread the OPORDER while you insert the Laura Disk into Drive 2. Only the first and second ID= lines are needed, e.g., ID 1= 3 Doe J. ID 2= Laura.

e. At the beginning of the third trial:

(1) Ask the subject, "Do you see the target?" Ask him to point to it because sometimes they mistake the purple/blue quarry near the center of the map for the target.

(2) Tell the subject, "there are three bridges, cross any one you like, however, the one referred to in the Operation Order is the middle bridge, near the center of the map.

4. Some possible problems:

a. TANK2:TANKER won't run. You receive a message that says volume is filled. This means that disk storage space is inadequate to accomodate the next subject. Insert a new System (Tank 1) Disk in Drive 1 or, using FILER, delete Demo data (.TEXT) from disk.

b. Ready light won't come on. Emphasize that the subject should never press the Fire Button unless the Ready Light is on. Remedy, press the $ key (4 and shift) to terminate game, then start new game. Note, the $ command can be used to terminate game, then start new game. Note, the $ command can be used to terminate a game for any reason, however all data for that trial will be lost, i.e., not recorded on disk. If it is ever necessary to terminate a game and start a new one, please make an ID entry to that effect on the next trial, e.g., no ready light last trial.

c. Cursor hangs up in an apparently unobstructed area. Sometimes the color green overshadows orange, therefore, areas that appear green are actually orange. Solution, ask the subject to back out and reroute.

d. Subject forgets to Advance Clock. After the first trial is completed, remind him to advance clock.

e. If S runs into mountain road east of bridge, say, "that is a bombed out mountain road, back out and try another route."

f. Student asks you, "What is this all about?" Answer, the Research Institute is interested in a large number of tests that might be related to tank crew performance. This is one of them. Our interest is in the test, not you as an individual. This is a research project, therefore, your name will remain anonymous and your performance today will have no bearing on your future as a tanker.

Instructions for Students

Purpose

This game was produced to help you learn how to operate the equipment in order to later play a game that was designed to simulate some conditions that you, as a Tank Commander, might encounter on a battlefield.

Equipment. In order to make sure that you understand the instructions, as the game is described, please point to each item when it is mentioned. The black box in front of you has five controls that you will be using. (Ignore the black Stick/Paddle switch.) The controls and displays perform the functions and are operated as described below.

Joystick. The black Joystick controls the speed and direction of the cursor (the large white cross (X)) which represents your tank on the TV screen. Simply push or pull the stick in the direction that you wish to go. The more pressure you apply to the stick, the faster the tank moves. If you release the stick it will return to the center position and the tank does not move, however, it is not in the parked mode, described below.

Toggle Switch. This chrome switch determines whether the tank is in the moving mode or the parked mode. To place the tank in the moving mode, flip the toggle switch to the right position. When the tank is in the moving mode the words "MOVING ENABLED" will appear at the bottom of the screen and you can move the tank. To place the tank in the parked mode, flip the toggle switch to the left position. Please flip the switch now. When the tank is in the parked mode the following statements will appear at the bottom of the screen: Fire, Stake, Time. When in the parked mode you can fire at the target, stake a position to which you might later want to return, or advance the clock to speed up the game. The Fire command does not appear at the present time for reasons that will be explained in a moment.

Buttons. The three red buttons can only be used when you are in the parked mode. They can be used to fire your own weapon, to place a stake at a position to which you want later to return, or to advance the computer clock when you are ready to start the attack but do not want to wait a long period of real time. The fire button is the top button (labeled 0). It can be pressed when you are in the parked mode (toggle switch to the left), to shoot at the target. The stake button is the middle button (labeled 1). It can be pressed to stake a position when you are in the parked mode (toggle switch to the left). This button can be used to help you plan positions for attack or in preparation of defensive or alternate positions. The time button is the bottom button (labeled 2). It is used to advance the clock. While in the parked mode, press the bottom button and then push the joystick forward until the TIME displayed at the bottom of the screen is set at the time you want. If you should advance the clock too far, just pull back on the stick and the clock will run backward until you have reached the desired time. If you wish to fire immediately after advancing the clock, you must first flip the toggle switch to the moving mode, then back to the parked mode.

Map. The map represents the terrain over and around which you will be moving, staking, and firing. The following conditions prevail:

1. WHITE Roads. The tank can go on the roads.

2. GREEN Hills and Trees. The tank can enter and travel around in the green areas.

3. BLACK Areas. The tank can enter the uncolored (black) areas.

4. BLUE Rivers. The tank cannot enter the blue rivers. You can only cross a river on a bridge. Bridges are located where the roads (white lines) cross the rivers (blue lines).

5. ORANGE Mountains. The tank cannot enter the orange areas.

6. PURPLE buildings, vehicles, or other structures. The tank cannot enter the purple areas. Purple is usually used to represent a target. In this demonstration game the target is represented by the large purple T.

TV Screen. Starting at the top left side of the screen you will find the following information displayed:

T LOSS. TARGET LINE-OF-SIGHT. An Orange light will appear whenever there is an unobstructed view line between the present position of your tank and the target. In this game, you can only shoot the target or be destroyed yourself when your tank is in line-of-sight to the target (T LOS light is on). It can be used to help prepare battle positions in advance. In this game your tank will not actually be destroyed, nevertheless, you should operate as though the target is an actual threat. As you can see the light is not on now because there is an orange mountain and a green hill between the tank and the target.

TARGET RANGE. The numbers displayed below the T LOS light tell you the distance (range) between your present tank position and the target. This range can be used for planning positions from which to attack the target. Read those numbers aloud.

B LOS. BRIDGE LINE-OF-SIGHT. A purple light appears underneath the word B LOS when there is a direct line of sight between your tank position and the bridge located near the center of the map. B LOS can be used for planning battle positions and staking alternate positions. Ask the student, "why is the light not on now?" (The student should answer, "because of the mountain between tank and bridge").

BRIDGE RANGE. The numbers displayed below the B LOS light tell you the distance (range) between your present tank position and the bridge located near the center of the map. This range can be used for planning positions for protecting the bridge. Read those numbers aloud.

STAKE. A white light appears beneath the word STAKE under the following conditions:

1. When you are in the parked mode and press the stake button (labeled "1").

2. When you are in the Moving Mode and cross a spot that has been previously staked. If you stop at that position and change to the parked mode the stake light will remain on.

READY. A green light will appear under the word READY when you are in position to fire. The light will appear <u>only</u> under the following conditions:

1. When you are in the parked mode.

2. When you are in line-of-sight to the target (the T LOS light is on).

3. When the distance to the target is within range of your own weapon.

4. After you have fired there is a five (5) second delay before you can fire again from that position. If you wish to fire again from that position, <u>DO NOT PRESS THE FIRE BUTTON UNTIL THE READY LIGHT IS ON</u>.

CIRCLE. At the lower left corner of the screen there is an outline of a circle with a line in it. The line points to the direction that the tank is headed in the moving mode. The line is meaningless in the parked mode.

TIME. At the bottom center of the screen, beneath the map, you will see time displayed. Time is constantly updated. For this demonstration, start time was set at 1730. The time displayed is computer time, which is three (3) times as fast as real time.

COMMANDS. The bottom line, underneath TIME, tells you which mode you are in and what you can do. You can be in one of the two modes, described above under "toggle switch."

MOVING MODE. When you are in the Moving Mode the statement "MOVING ENABLED" will appear. In this mode you may move the tank in any direction and at whatever speed you desire. Maximum speed, in relation to the map, is approximately 90 kilometers per hour.

PARKED MODE. When in this mode the statement "COMMANDS: TIME, STAKE, FIRE" appears. The FIRE command will only appear when READY light is on (in T LOS and within own weapon range). When you have advanced the clock and are ready to fire, you must flip the Toggle Switch to Moving Mode and then back to Parked Mode.

Instructions for the Demonstration Game

The test administrator shall read the following instructions to the student.

"Your tank is in the assembly area at the lower left corner of the screen. Place it in the moving mode, push the joystick forward and watch your tank move north. In order to help you understand how to play the game, please do the following:

1. Go north into the green wooded hilly area, avoiding the orange mountain, and place a stake so that you can later destroy or protect the bridge, if you are so ordered. As you are proceeding north, notice that the B LOS light comes on as you pass the orange mountain and the bridge comes into sight.

2. Once you have placed a stake in the green area, proceed to the bridge and cross it. Notice that as you approach the bridge the T LOS light comes on when the target is in sight.

3. After you have crossed the bridge go to the green area and place a stake in a position from which you could shoot the target.

4. Now proceed to any position of your choice in a green area from which you can fire (remember, you must have line-of-sight to the target).

5. Let's assume that your orders are to fire at 1800. Once you are in position to fire (Ready Light is on), instead of waiting for real time to pass, advance the clock. The clock is advanced, while in the parked mode, by firmly pressing the bottom, number two, button and then pushing the joystick forward. As you push the joystick forward, you will see the time displayed at the bottom of the screen speed up. Stop when the clock reaches 1800. If you go past 1800, just reverse the joystick and run the time back. After you have advanced the clock, flip the toggle switch to the right and then quickly back again, so that you can fire.

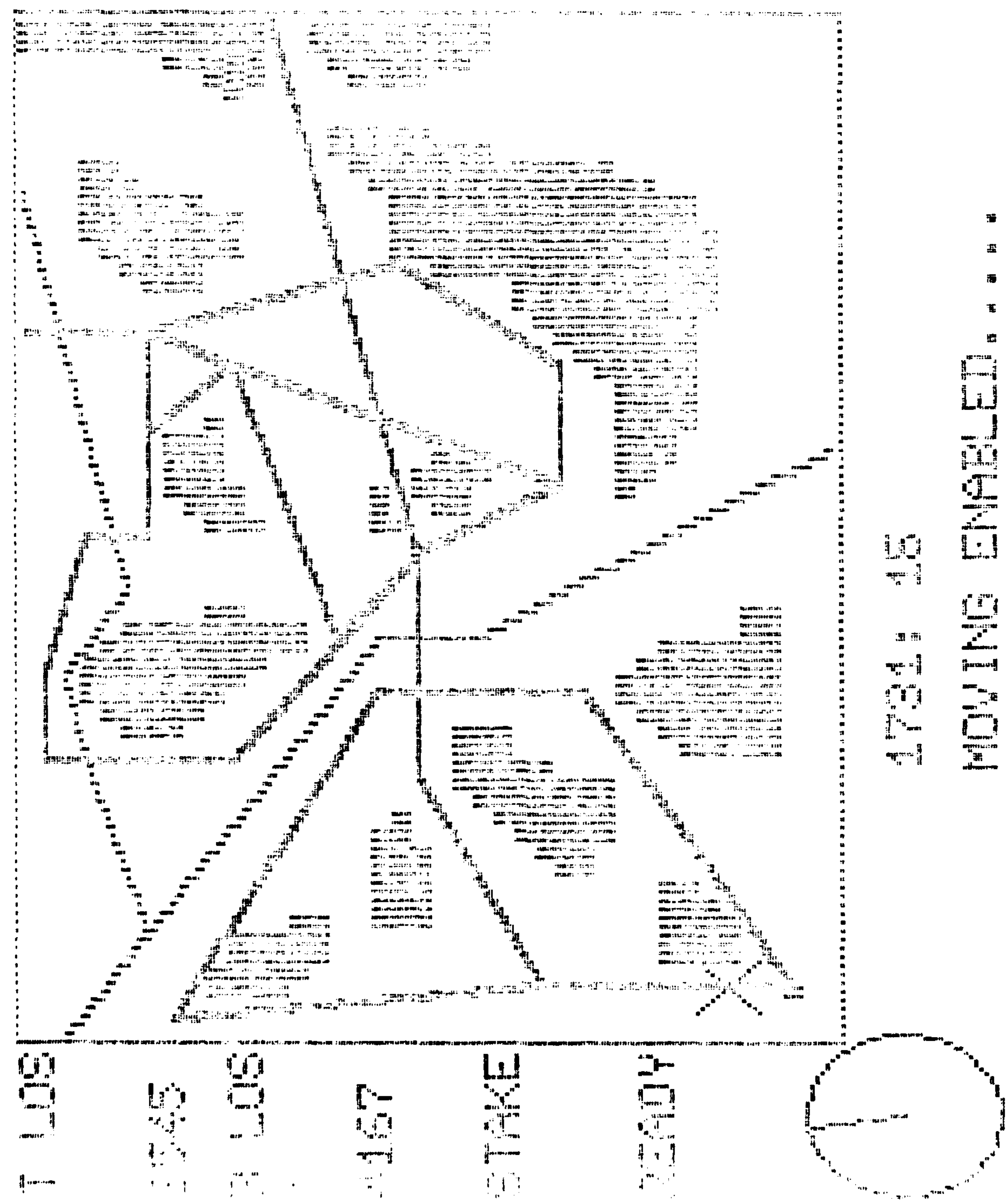
6. Now you are now ready to fire because you are in the parked mode and the ready light is on. When told to do so, press the top, number zero, button to fire your weapon. When you fire you will see a line flash between your tank and the target. You will always have to fire more than once to destroy the target, even though you have apparently hit it. Ready-Fire.

7. After you have fired the first time you may wait at that same place and fire again, after the ready light appears, or you may move to another position and try again. In any event, always wait

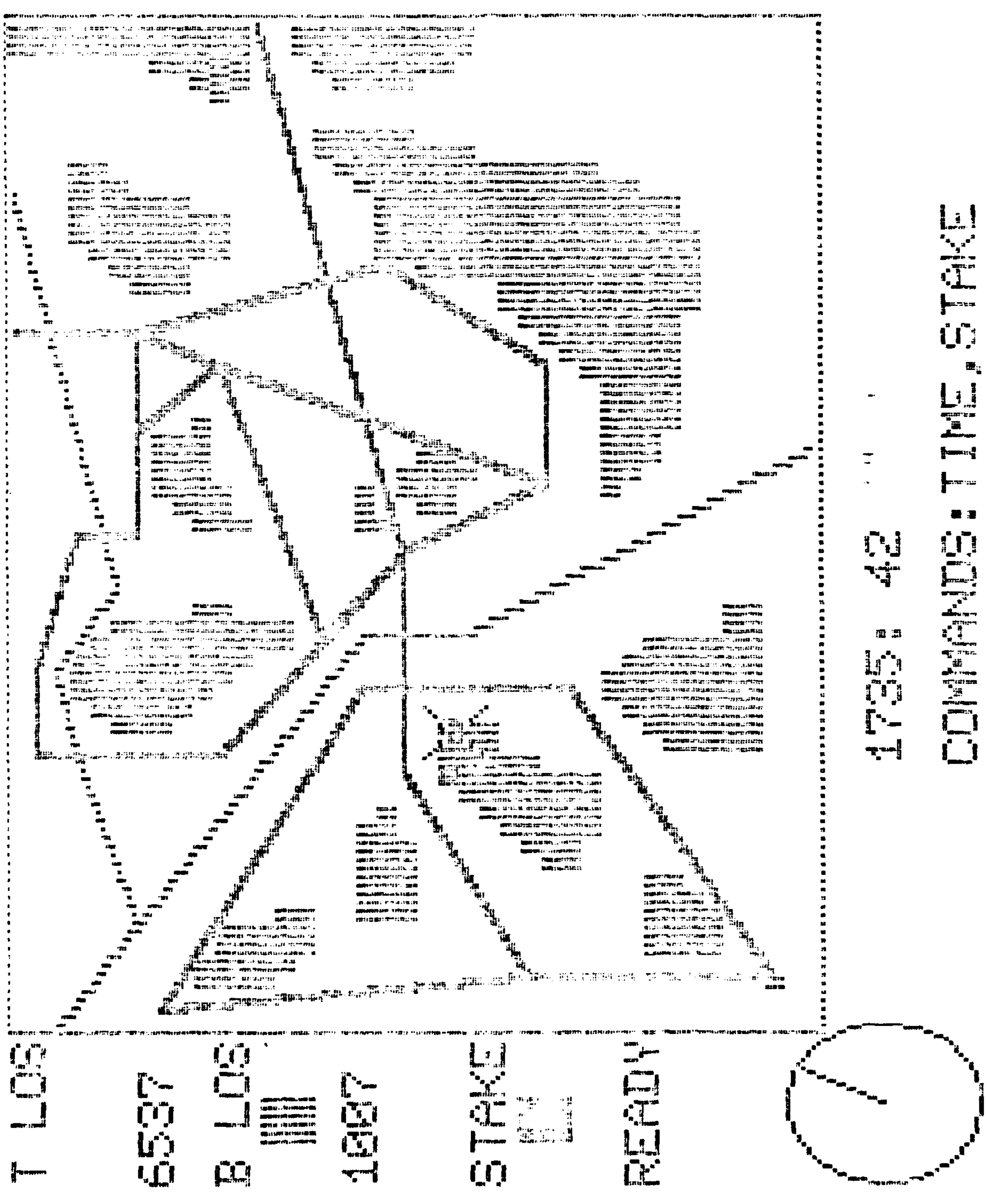
for the Ready Light. Continue to fire until the target is destroyed (there will be sound effects and shooting stars). The game ends when the target is destroyed.

Appendix D

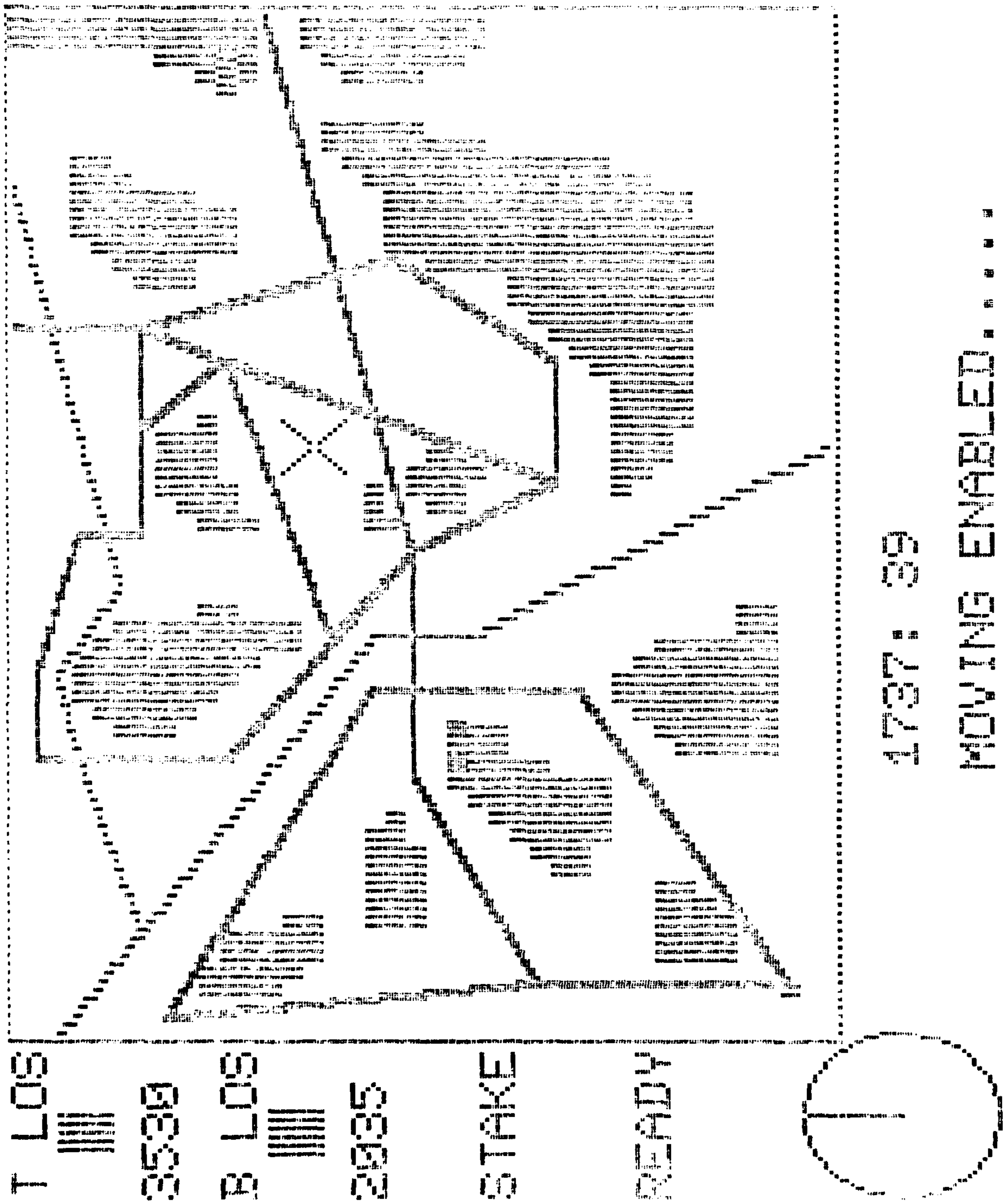
Examples of Computer Maps

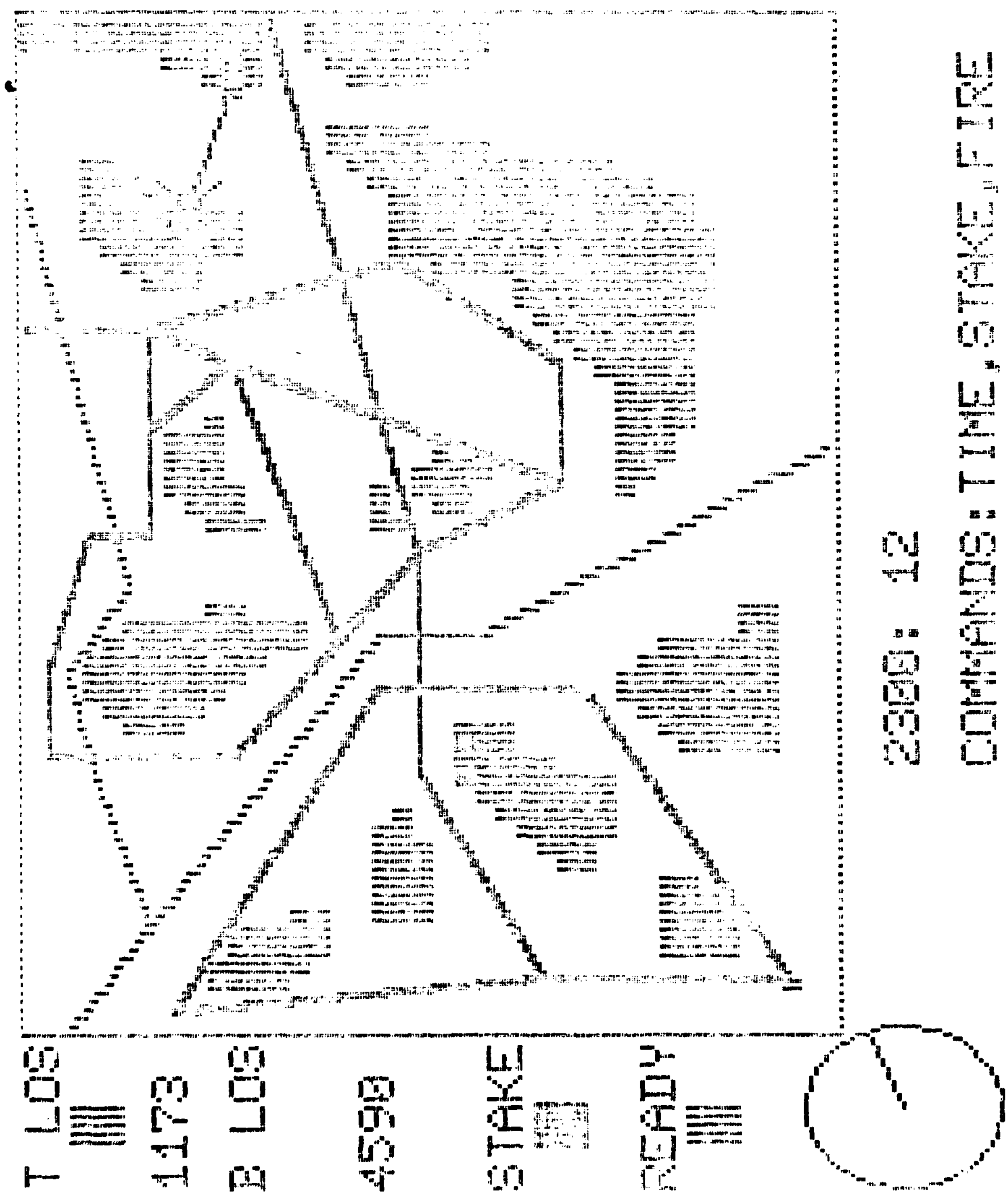


T LOS
745
LOS
167
STAKE
1731: 15
MOVING ENABLED....



T LOS
6537
B LOS
1987
STAKE
READY
1735: 42
COMMANDS: TIME,STAKE


T LOS
3539
B LOS
2335
STAKE
READY
1737: 39
MOVING ENABLED....


T LOS
1173
B LOS
4598
STAKE
READY
2300: 12
COMMANDS: TIME, STAKE, FIRE

Appendix E

The Operational Order (OPORDER) Used for the TC Video Game

Operation Order

You are a Tank Commander of an M60A1 with the 3d PLT, Team B, of Task Force 2-10 Armor. The Team is currently in its assembly area. Local time is 1730. Team B will be leading a movement to contact and will over watch the attack to seize an objective on the right edge of your sketch map, shown as the purple rectangle on the screen. Your Platoon Leader has given you the following mission:

1. Move to the hill over watching the objective and at 2300 hrs engage (suppress) the ATGM (purple color). You are to arrive at that position no later than 1800. You will have plenty of daylight time to prepare for the night time attack, therefore, after you have completed the preparation of positions, advance the clock to 2300 and fire.

2. Your Platoon Leader also directs you to move out early so you can prepare positions on the west (left) side of the river. On order you will move to those positions to protect the bridge which is on the Brigade Main Supply Route. The Lieutenant expects to occupy these positions and defend the bridge under cover of darkness after the 2300 attack on the primary target, an anti-tank guided missile (ATGM). Remember, you should prepare the positions to defend the bridge before proceeding to the primary target.

3. The enemy has AT4s (2000m) and Hind Es equipped with AT6s (4000m). He can adjust artillery on any target within one minute. Range of your own weapon is 2000m.

Appendix F

Paper-and-Pencil Instructions and Procedures

Paper-and-Pencil Instructions

Please leave the tests face down until I tell you to turn them over. This session is paper-and-pencil testing from an experimental battery; therefore, there are no final norms and we are not interested in evaluating your IQ or individual abilities. Please take the test seriously though, and do your best because we are interested in how they correlate with experience on the video games. If you have any questions about any of this, you may ask me later.

When I ask you to turn the tests over, please print your name on the top sheet: Last Name, comma, First Name, comma, Middle Name. Also, write down your Social Security number. Do this on the first sheet only. There are places on each test for your name. Leave them blank. We ask this to help maintain anonymity when we read the results. We will assign a code number to your name which will appear on the top sheet only. Later, your code number will be filled in on each test.

On the first page of each test, there are instructions which you are to read and examples which may help you understand the tasks. This page is followed by the actual test.

Your first test is to find hidden figures to match a model. When I say TURN, turn to the first page of the Hidden Patterns test and study the instructions and examples carefully. Then, when I say TURN after you have read the instructions, turn to the first test page and start working. At the bottom of each page it tells you to STOP or GO ON TO THE NEXT PAGE. Follow the directions at the bottom of the page. And when I say STOP, stop at once and wait for me to tell you to go on. Most of the tests have two parts, and some of them have more pages than you can finish. So when I say, TURN TO PART TWO, make sure that you turn to Part 2. Some of the tests are printed on both sides of the paper; so don't skip any questions. You will have time to breathe between tests and a short break halfway through.

Are there any questions on the procedure?... If you have any trouble during the tests, signal for help.

Ready...."TURN"

Paper-and-Pencil Procedures

Test 1 Instructions: 2 minutes

"STOP... Remember to mark each figure in the row. Mark an X if it includes the model, an 0 if it doesn't. READY... TURN"

Part 1 3 minutes

"STOP... READY... TURN TO PAGE 2"

Part 2 3 minutes

"STOP... The next test is Grammatical Reasoning,

"READY... TURN"

Test 2 Instructions 30 seconds

"READY... TURN"

Test 1 minute

"STOP... The next test is planning the shortest route along a map. Read the instructions carefully. They give you the key to the symbols on the maps. You'll be lost if you don't know the symbols and the rules.

READY... TURN"

Test 3 Instructions 4 minutes

Part 1 3 minutes

"STOP... TURN TO PART 2"

Part 2 3 minutes

"STOP... Next is a paper folding test.

READY... TURN"

Test 4 Instructions 4 minutes

"STOP... READY... TURN"

Part 1 3 minutes

"STOP... READY... TURN TO PART 2"

Part 2 3 minutes

"STOP... The next test is an addition test. You have five minutes to complete it.

READY... TURN... BEGIN"

Test 5 TEST 5 minutes

"STOP... That's it. I have a handout for the video and computer games. The rest of the experiment is a lot more fun."

Appendix G

Video Game Experience Questionnaire

NAME (PLEASE PRINT) ________________________________

PHONE ____________ ADDRESS __________________________

1) Please estimate how many hours you have played the ATARI home video games (not the ones at the airport for example).

2) How many different kinds of ATARI home video games have you played? _____

How many different cassettes? ____

3) Which particular game have you played extensively (if you don't remember please describe briefly)

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.

4) Do you or your family own an ATARI home video game? ____

5) How would you rate yourself relative to other ATARI players (make an X at the appropriate point).

poor ________________________ great

6) Have you played video games other than the ATARI home video games extensively? ____ If so, list below.

PT: 5541

Appendix H

Instruction and Scoring for Combat (ACM)

Instruction for Combat

The Game

You are looking down on the action. You are the good guy in the light-colored plane. Travelling across your screen is an enemy dark-colored plane. You must hit him as many times as you can in the space of two minutes and 16 seconds.

Using the Controller

Hold you joystick with the red button to your upper left, toward the television screen. The speed of your jet is controlled by moving your joystick from the forward position (fastest) to the back position (slowest). Right and left turns are controlled by moving your joystick from left to right.

Fire your missile by pushing the red button on your joystick controller. A direct hit on your target will change his course requiring you to maneuver for the next shot.

When your time is up and the game is over, your plane will no longer respond to your Joystick. To start again, simply pull down on the reset lever (far right of game console) and begin.

Combat Score Sheet

Name	Start Time	Stop Time
Trial	Score	
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Appendix I

Instructions and Scoring for Touch-Me

Instructions for Touch-Me

In Touch-Me you must repeat the lights and sounds the computer gives you as the length of the series grows with each "round."

To Begin

Make sure the game is turned on (switch at lower left) and set for game 1 (switch at lower right): "0 0" will appear on the readout. Press the START button (top far left) once. The computer will display alternately "08, 16, 32, 99" repeatedly. These are skill levels corresponding to the number of lights/tones in the series you will be attempting to repeat. When the display comes around to the skill level "99", press the yellow button marked "SKILL" (far right). You are now ready to begin.

Press the START button again. The computer will pick one of the four colored areas causing that area's light to go on and playing the tone (Note #1) associated with that color. You must press the colored button in that area to repeat Note #1. The computer will then play Note #1 and add a second note/light. You must repeat both Notes #1 and #2 in the order the computer gave them to you. Each time you repeat what the computer has given you, it will add one more note to the series. Continue repeating the ever-increasing series until you reach the number of your skill level. The number of notes in each "round" is displayed up above.

If you push the wrong color or hesitate too long, the computer will give you a "raspberry". It will then repeat the last series for you and you may try again. Three misses before reaching your skill level and the game is over. Once the game is over, when you have missed three in a row, set the skill level to 99 (press the START button once and stop the display at 99 by pressing the SKILL button). You are now ready to begin the test.

Touch-Me Score Sheet

Name	Start Time	Stop Time

Trial	Score
1	
2	
3	
4	
5	
6	
"	
"	

Appendix J

Rating Scales For Criterion Measurement

BNCOC Cadre Evaluation of Student Performance

The first sentence, for each of the five items below, states the skill area to be rated. There are four descriptors (numbered zero to three) for each skill area. Check the one, based upon your experience, that applies to the following student. If you have not observed this student's behavior directly, please make your best estimate.

STUDENT NAME:

1. A good Tank Commander does not expose his tank to the threat of enemy fire anymore than absolutely necessary. This Student:

0 - skylines or goes right up the middle, pays no attention to enemy threat

1 - avoids obvious hazards but does not take advantage of positive terrain features

2 - usually follows tree lines and terrain contours, but in operational situations, occasionally forgets

3 - second nature to use terrain to his advantage and uses it effectively.

2. A good Tank Commander fires from good battle positions, primary and alternate. This Student:

0 - selects inferior primary battle position and never thinks about alternate battle positions

1 - picks adequate primary position but does not plan good alternate positions

2 - selects good primary position and occasionally good alternate positions also

3 - almost invariably chooses good primary position and alternate positions.

3. A good Tank Commander moves after firing. This Student:

0 - never moves after firing

1 - sometimes moves but usually too late

2 - usually moves but sometimes too late

3 - always moves in time to afford low likelihood of return fire.

4. A good Tank Commander conforms to training and tactical doctrine. This Student:

0 - does not know what training and doctrine require

1 - is familiar with doctrine but not able to execute in the field

2 - understands doctrine and executes appropriately, but not always on time

3 - knows doctrine and implements it with authority.

5. A good Tank Commander manages his resources (crew, fuel, ammunition, provisions, etc.) by planning ahead. This Student:

0 - never thinks past the immediate moment

1 - remembers major items but calculates incorrectly, or neglects to take corrective actions

2 - usually manages major resoures well, occasionally forgets minor items

3 - plans ahead and pays attention to details.

Overall Evaluation of This Student

Please rate this student relative to your general experience with other past tank commander students. Check on of the following categories.

top 10% (90-100%)

next to top 25% (65-90%)

middle 30% (35-65%)

next to bottom 25% (10-35%)

bottom 10% (0-10%)

If for any reason you do not think that this evaluation of the student's performance reflects his real potential, please state what these reasons are.

BNCOC Class Ranking

Please rate all students in the course from most proficient (number 1) to least proficient, on their overall performance in the BNCOC course.

NAME

1) ________

2) ________

3) ________

4) ________

5) ________

6) ________

7) ________

8) ________

9) ________

10) ________

11) ________

12) ________

Statement Used for Peer Rating

You have been ordered to undertake a dangerous mission where your probability of survival is poor and your likelihood of success is dependent on the abilities of your tank commander. You are assigned to the loader position. Which three individuals in your class would you most like to see as your tank commander. Rank the best in your opinion first, then the second best, then the third best.

1.

2.

3.

Appendix K

Correlation Matrices for the Three BNCOC Classes

Correlation Matrix for BNCOC M1 (N=7)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	$\bar{X}$	SD
1. Rankit	--	.96**	.95**	-.62	-.41	-.64	-.48	-.07	-.29	.29	-.59	.53	.43	.55	.03	.29	0.0	0.9
2. BARS Summed		--	.97**	-.77*	-.59	-.73*	-.65	.10	-.36	.49	-.69	.60	.21	.59	.02	.46	8.4	3.4
3. Percent Standing			--	-.79*	-.55	-.79*	-.61	.05	-.40	.45	-.69	.61	.15	.61	.02	.38	37.3	27.2
4. LOS				--	.71*	.85**	.64	-.13	.37	-.66	.54	-.63	.33	-.70	.03	-.63	296.5	77.1
5. LOSX					--	.85**	.96**	-.45	.74*	-.29	.85**	-.13	.41	-.15	.28	-.31	951.5	237.1
6. MAZ						--	.79*	-.16	.72*	-.23	.75*	-.29	.23	-.42	.28	-.28	109.1	56.7
7. LOSP							--	-.45	.67	-.26	.94**	-.12	.32	-.06	.33	-.32	63.0	20.6
8. AVE								--	-.50	.30	-.39	.33	-.33	.15	.43	.08	2.0	0.1
9. # of Stakes									--	.22	.58	.06	.14	-.10	.29	.06	11.9	4.8
10. Mathematics										--	-.17	.80*	-.43	.64	.45	.72*	31.0	12.7
11. Grammatical Reasoning											--	-.09	.19	.02	.21	-.09	6.1	3.2
12. Hidden Patterns												--	-.02	.92**	.59	.61	147.9	84.0
13. Map Planning													--	.07	-.05	-.15	10.8	6.8
14. Paper Folding														--	.40	.61	6.9	5.4
15. ACM															--	-.13	9.3	2.6
16. Touch-Me																--	13.4	4.2

*$p < .05$. **$p < .01$

Correlation Matrix for BNCOC AI (N=12)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	$\overline{X}$	SD
1. Rankit	--	.64*	.57*	-.65*	-.61*	-.68**	-.65*	-.59*	-.52	.32	-.27	.63*	.13	.73*	.42	.32	0.0	0.9
2. BARS Summed		--	.76**	-.43	-.29	-.37	-.33	-.56*	-.78**	.18	-.47	.54	-.16	.35	.07	.35	8.2	1.9
3. Percent Standing			--	-.50	-.39	-.43	-.42	-.59*	-.51	.10	-.30	.58*	.22	.59*	.18	.56*	60.2	14.1
4. LOS				--	.96**	.95**	.97**	.36	.85**	-.32	-.04	-.92**	-.40	-.82**	-.45	-.72**	265.8	197.0
5. LOSX					--	.93**	1.00**	.32	.78**	-.31	.02	-.77**	-.38	-.78**	-.33	-.65*	1,101.5	1,058.0
6. HAZ						--	.95**	.38	.71**	-.17	.00	-.88**	-.41	-.86**	-.57*	-.59*	130.2	147.0
7. LOSP							--	.37	.76**	-.29	.05	-.79**	-.38	-.81**	-.39	-.63*	73.6	76.6
8. AVE								--	-.02	-.18	.29	-.37	-.28	-.39	-.40	-.01	1.9	0.3
9. # of Stakes									--	-.22	.30	-.87**	.03	-.57*	-.24	-.76**	12.9	5.9
10. Mathematics										--	.42	.24	.38	.13	-.26	.27	35.5	6.8
11. Grammatical Reasoning											--	.13	.43	-.07	.04	.22	6.1	3.7
12. Hidden Patterns												--	.37	.79**	.60*	.73**	130.2	67.4
13. Map Planning													--	.40	.34	.19	12.6	7.2
14. Paper Folding														--	.65*	.58*	7.6	5.4
15. ACM															--	.07	8.6	2.5
16. Touch-Me																--	15.1	5.3

*$p < .05$. **$p < .01$

Correlation Matrix for BNCOC A3 (N=11)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	$\bar{X}$	SD
1. Rankit	--	.94**	.91**	.25	-.15	-.18	-.34	-.47	.31	.18	-.16	.08	.64*	.43	.69*	.61*	.79**	0.0	0.9
2. BARS Summed		--	.94**	.15	-.21	-.26	-.39	-.50	.34	.36	-.08	.15	.54	.16	.77**	.57	.82**	9.5	4.0
3. Percent Standing			--	.10	-.30	-.37	-.48	-.55	.35	.43	-.22	.11	.69*	.17	.63*	.58*	.73**	65.0	27.2
4. LOS				--	.88*	.79**	.77**	.13	.17	-.44	-.10	-.15	.19	.35	.14	.27	.17	232.5	98.2
5. LOSX					--	.93**	.98**	.29	-.09	-.46	.04	-.11	-.10	.15	-.02	.11	-.20	1,105.0	534.0
6. HAZ						--	.92**	.28	-.13	-.46	.19	-.10	-.13	.30	.06	.14	-.27	111.1	59.6
7. LOSP							--	.35	-.17	-.48	.13	-.07	-.23	.05	-.12	.04	-.38	65.8	36.1
8. AVE								--	-.32	-.25	-.07	.32	-.34	.26	-.55	-.47	-.26	1.8	0.4
9. # of Stakes									--	-.10	.22	-.14	-.00	-.22	.08	.25	.22	13.7	8.3
10. Mathematics										--	-.13	.02	.48	-.26	.14	.02	.29	40.9	10.3
11. Grammatical Reasoning											--	.41	-.27	-.23	.24	.43	-.37	4.0	1.8
12. Hidden Pattern												--	-.13	.03	.32	.47	-.22	167.8	51.1
13. Map Planning													--	.46	.19	.41	.50	15.0	5.0
14. Paper Folding														--	.08	.19	.27	2.2	5.5
15. ACM															--	.71**	.41	9.8	4.0
16. Touch-Me																--	.05	12.5	5.6
17. Peer Rating																	--	1.8	5.9

*$p < .05$. **$p < .01$